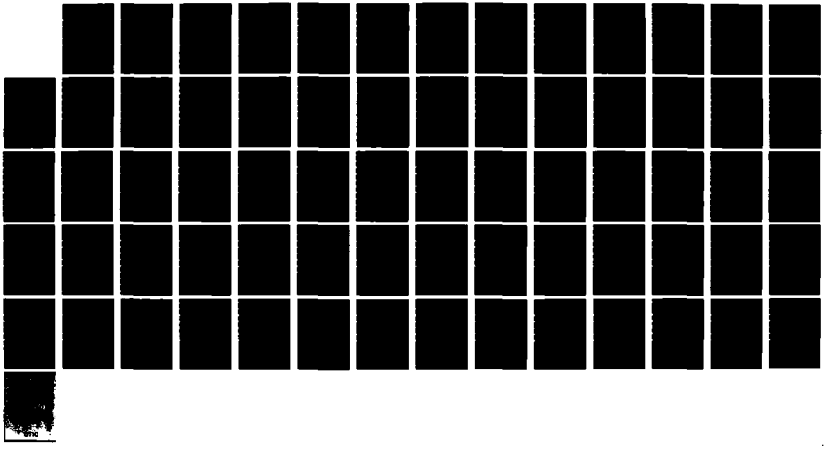
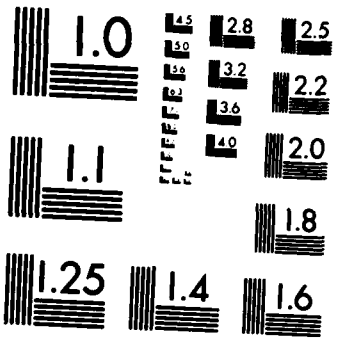
AD-A129 781 RECENT PAPERS IN PARAMETRIC MODELLING OF TIME SERIES (U) RHODE ISLAND UNIV KINGSTON DEPT OF ELECTRICAL ENGINEERING L L SCHARF ET AL. APR 83 N00014-82-K-0300 1/1

UNCLASSIFIED F/G 17/2 NL





RECENT PAPERS IN PARAMETRIC MODELLING OF TIME SERIES

Louis L. Scharf
Claude J. Gueguen*
Jean-Pierre Dugre

Department of Electrical Engineering
University of Rhode Island
Kingston, Rhode Island 02881

*Ecole National Superieure des Telecommunications
46 rue Barrault
Paris 75634

Contents:

Prepared for

OFFICE OF NAVAL RESEARCH (Code 411 SP)
Statistics and Probability Branch
Arlington, Virginia 22217
under Contract NOOO14-82-K-0300
L. L. Scharf, Principal Investigator

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A129 781	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Recent Papers in Parametric Modelling of Time Series		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Louis L. Scharf Claude J. Gueguen Jean-Pierre Dugre		8. CONTRACT OR GRANT NUMBER(s) N00014-82-K-0300
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Electrical Engineering University of Rhode Island Kingston, Rhode Island 02881		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 411 SP) Department of the Navy Arlington, Virginia 22217		12. REPORT DATE April 1983 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Moving average models, autoregressive models, linear transformations, linear filters, spectrum analysis, fast algorithms

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In the set of reports we discuss a variety of modern topics in signal modelling, parameter estimation, and spectrum analysis. Throughout, the emphasis is on linear transformations of random data, and the insight this gives into topics as diverse as spectrum analysis, filtering, and modelling.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

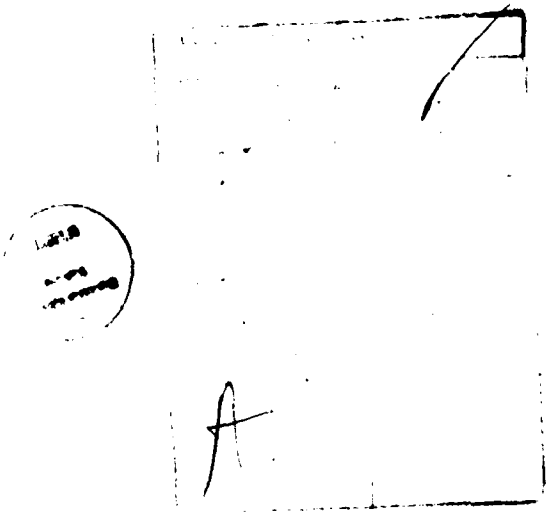
EXACT MAXIMUM LIKELIHOOD IDENTIFICATION OF ARMA MODELS:

A SIGNAL PROCESSING PERSPECTIVE

Claude Gueguen
Louis L. Scharf

Presented at:

1st European Signal Processing Conference
Dept. of Elec. Engr., Swiss Federal Inst. of Tech.
Lausanne, Switzerland
September 1980



Sept. 16-19 1980 | Lausanne Switzerland | First European Signal Processing Conference | EUSIPCO-80
Dept. of Elec. Engr.
Swiss Federal Inst. of Tech.
Lausanne, Switzerland

EXACT MAXIMUM LIKELIHOOD IDENTIFICATION OF ARMA MODELS: A SIGNAL PROCESSING PERSPECTIVE[1]

Claude Gueguen
Ecole Nationale Superieure de Telecommunications
46 rue Barrault 75634 Paris
(on leave at Colorado State University, Fort Collins, CO 80523)

Louis L. Scharf
Electrical Engineering Department, Colorado State University
Fort Collins, CO 80523

Abstract

Autoregressive-moving average (ARMA) models, and their autoregressive (AR) counterparts, are useful approximants to the kinds of random processes commonly encountered in discrete-time signal processing applications. Such models may be used to compress data in low bit-rate information transmission, improve frequency resolution in spectrum analysis, and to forecast in economic, meteorological, and other time series.

In this paper we discuss several aspects of the maximum likelihood theory of parameter identification in ARMA and AR models. We highlight the initial condition problem encountered when identifying ARMA or AR models from finite data records and propose several methods for computing exact and approximate likelihood. Several new interpretations are given for the innovation representation of an ARMA process. Computationally efficient lattice and fast Kalman filters are proposed for the computation of exact likelihood.

1. INTRODUCTION

The random processes encountered in signal processing applications are typically lowpass or bandpass processes in which redundancy is high. This means finite-dimensional models may often be used to approximate the second-order properties of the processes. The dominant motivations for using finite-dimensional models are (1) they provide a systematic framework for deriving data compression and frequency resolution improving algorithms, and (2) they become predictor formulae for event forecasting.

The problems of data compression, resolution improvement, and forecasting are "solved", so to speak, by identifying a parametric model that either infinitely extends a data correlation sequence or matches the data, itself, in a least squares or maximum likelihood sense.

The catch in all of this is model selection and parameter identification within the model. Autoregressive (AR) models are by far the simplest parametric models to identify. Exact maximum likelihood theory leads to nonlinear methods, even in the AR case. However, the best AR predictor is finite-memory and linear identification procedures involving Toeplitz normal equations or non-Toeplitz Yule-Walker equations are routinely used.

AR models suffer the defect that spectral zeros are not easily modeled with low-order schemes. Couple to this defect the fact that sample-data versions of rational continuous-time processes are autoregressive moving average (ARMA) [11], and we have strong motivation for identifying the more general ARMA models.

Traditionally the emphasis in identification of ARMA models has been on approximate representations (such as "long ARs") that lead to linear identification procedures. However, more recently there has been a flurry of activity in exact maximum likelihood formulations and nonlinear optimization procedures. Representative recent offerings include papers by Akaike [1], Newbold [2], Osborne [3], Harvey and Phillips [4], Ansley [5], Pearlman [6], and Jones [7]. Akaike [1] has advocated the use of a Markovian representation for an ARMA process, which is essentially the standard form state space model well-known to engineers. Jones [7] has used this representation to formulate exact likelihood equations for ARMA processes observed in white noise. He advocates state space models for the calculation of exact likelihood with or without missing observations. Newbold [2] generalizes the exact likelihood results of Osborne for MA processes, and of Box and Jenkins [8] for AR and MA processes, to include ARMA processes. Harvey and Phillips [4] advocate the Kalman filter as a recursive technique for computing exact likelihood and reference Schweppe [9] as perhaps the first investigator to write exact likelihood in terms of prediction errors

[1]This work supported by the Army Research Office, Research Triangle Park, NC 27709, under contract DAAG29-79-C-0176 and by the Office of Naval Research, Statistics and Probability Branch, Arlington, VA 27740, under contract N00014-75-C-0518.

or "innovations". Pearlman [6] discusses the fast algorithm of Morf, Sidhu, and Kailath [10] for updating the Kalman gain, and compares it to the algorithm of Ansley [5].

Why exact likelihood and maximum likelihood theory? Perhaps the most convincing argument in favor of an exact maximum likelihood formalism for identifying ARMA models is that it gives one a basis from which to approximate. We return to this point in Section 5. Add to this argument the success of authors like Jones [7] with maximum likelihood identification of low order (e.g. $p \leq 3$) ARMA models. Whether or not exact maximum likelihood becomes a standardized identification procedure will depend in large part on our ability to efficiently compute likelihood and to make good parameter adjustments using nonlinear optimization procedures.

Paper Outline: In this paper we begin with a general discussion of maximum likelihood (ML) theory for identifying ARMA models of normal time series. A modal decomposition of the correlation matrix for ARMA processes is derived and placed in context with Anderson's work [12] on correlation matrix identification when the matrix is a linear combination of known matrices. The identification equations that result from ML theory are matrix versions of the Aigrain-Williams equations [13].

We next derive a linear time-invariant predictor formulation of the likelihood function based on a standard form or Markovian state space representation for an ARMA time series. In this formulation the likelihood function is an infinite-dimensional average over a noncountably infinite collection of conditional likelihood functions. The conditioning is on an initial condition (or state) vector. The values of this representation are these: (1) special initial condition assumptions suggest themselves for approximating the exact likelihood function (two of the most popular techniques associated with the covariance method of linear prediction are easily interpreted in terms of initial condition assumptions) and (2) an interesting ML procedure for ARMA parameters and initial conditions arises as a potentially useful method of approximating the exact likelihood.

We use the results of our linear time-invariant predictor formulation, together with variations on the Bayesian tricks used in Box and Jenkins [8] and Scharf and Nolte [14] to derive exact likelihood for ARMA processes. The prediction residuals from a linear time-invariant predictor are used in conjunction with a least squares estimate of the initial state based on the observed data record. The results include those of Newbold [2], Osborne [3], and Box and Jenkins [8] as special cases.

The next item of business is a time-varying (innovations) representation of the likelihood function in which the Kalman gain arises as the principal computational problem. We show how this gain may be obtained by triangularizing an (NxN) correlation matrix. This provides a fast Kalman predictor of the Morf, Sidhu, Kailath [10] variety. From the perspective of filtering or time series analysis the innovations representation of an ARMA process provides a zero-initial condition, time-varying linear filter representation of a stationary process. From a batch data processing point of view the innovations representation provides a sequential triangularization of the inverse correlation matrix. Finally, viewed from a probabilistic perspective, the innovations representation solves the Chapman-Kolmogorov equation that arises in connection with our averaged linear time-invariant representation for the likelihood function.

The innovations representation is shown to be equivalent to a representation in which data is generated as the output of a sequence of AR filters of order 1,2,... . This leads to an AR lattice for computing exact likelihood for ARMA models.

An efficient fixed point algorithm for computing the Kalman gain is discussed and computational requirements are compared with those of Ansley [5] and Morf, Sidhu, and Kailath [10].

2. MAXIMUM LIKELIHOOD THEORY FOR ARMA PROCESSES

Let $Y = (y_0, \ldots, y_N)'$ denote a finite version of a wide-sense stationary Gaussian sequence $\{y_t\}$. Assume the mean-value sequence is identically zero. Denote the correlation sequence by $\{r_p\}$. The correlation matrix for Y is the non-negative definite Toeplitz matrix

$$R_N = \begin{bmatrix} r_0 & r_1 & \cdots & r_N \\ r_1 & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \\ \cdot & & & r_1 \\ r_N & & & r_0 \end{bmatrix}$$

The likelihood function for a realization of Y is

$$\ell_N(Y) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|R_N| - \frac{1}{2}Y'R_N^{-1}Y \qquad (1)$$

We make no notational distinction between random variables and realizations of them, relying instead on context to make the meaning clear.

A typical ML inference problem is to maximize the likelihood function with respect to the correlation matrix R and call the resulting "estimate" the ML estimate of R. The result is

$$\hat{R}_N = \arg\max_{R_N} \ell(Y)$$
$$= C \; ; \; C = YY'$$

Thus, without prior parametrization of R_N to

reflect the fact that it must be Toeplitz and related to an ARMA sequence, ML theory leads to an inefficient, inconsistent, non-Toeplitz estimate of R. The corresponding spectrum estimate is the periodogram, a notoriously bad estimate.

Modal Decomposition: The correlation sequence of an ARMA (p,p) process (p poles and p zeros) may be written

$$r_n = \sum_{m=1}^{p} A_m \rho_m^{|n|}$$

We will call $\{\rho_m^{|n|}\}$ the m^{th} mode of the process and A_m the corresponding residue. Each ρ_m is a complex zero of the polynomial $A(z) = 1 + a_1 z^{-1} + \ldots + a_p z^{-p}$ in the ARMA representation

$$y_t + a_1 y_{t-1} + \ldots + a_p y_{t-p} = u_t + b_1 u_{t-1} + \ldots + b_p u_{t-p}$$
$$u_t : \text{i.i.d } N(0,\sigma^2) \text{ r.v.s.} \qquad (2)$$

We assume the zeros of A(z) and $B(z) = 1 + b_1 z^{-1} + \ldots + b_p z^{-p}$ lie inside the unit circle. The resulting process is stable and minimum phase.

The covariance matrix R_N may now be written as a linear combination of symmetric, linearly-independent, Toeplitz modal forms:

$$R_N = \sum_{m=1}^{p} A_m G_m$$

$$G_m = \begin{bmatrix} 1 & \rho_m & \cdots & \rho_m^{N-1} \\ \rho_m & & & \\ \vdots & & & \\ \rho_m^{N-1} & & & 1 \end{bmatrix}$$

An alternative representation is [12]

$$R_N = \sum_{m=1}^{p} A_m \sum_{t=0}^{N-1} \rho_m^t F_t = \sum_{t=0}^{N-1} r_t F_t$$

$$F_t = \begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 1 & & & & & \\ 0 & & & & & 1 \\ \vdots & & & & & \vdots \\ 0 & \cdots & 1 & & & 0 \end{bmatrix}$$

(t^{th} diagonal)

These representations correct a basic defect in the original formulation by forcing the ML estimate of R to be symmetric and Toeplitz.

Write

$$\hat{R}_N = \sum_{m=1}^{p} \hat{A}_m \hat{G}_m$$

where

$$\{A_m,G_m\}_1^p = \arg \max_{\{A_m,\rho_m\}_1^p} -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|R_N| - \frac{1}{2}\operatorname{tr} R_N^{-1} C$$

The resulting nonlinear ML equations are

$$\text{for poles } \rho_m: \operatorname{tr} \hat{R}_N^{-1}[D_m \hat{R}_N^{-1} C - D_m] = 0 \quad (m=1,2,\ldots,p)$$

$$\text{for residues } A_m: \operatorname{tr} \hat{R}_N^{-1}[\hat{G}_m \hat{R}_N^{-1} C - \hat{G}_m] = 0 \quad (m=1,2,\ldots,p)$$

$$\text{constraints: } \hat{R}_N = \sum_{m=1}^{p} \hat{A}_m \hat{G}_m \quad \hat{G}_m = \{\hat{\rho}_m^{|i-j|}\}$$
$$\hat{D}_m = \{|i-j|\hat{\rho}_m^{|i-j|}\} \qquad (3)$$

Here is how these equations may be used. For any choice of residue-pole pairs $\{A_m,\rho_m\}_1^p$ objective functions of the form $\mu(A_m,\rho_m) = \operatorname{tr} R_N^{-1}[G_m R_N^{-1} C - G_m]$, m=1,2,...,p, may be formed. These functions have zeros at $\{\hat{A}_m,\hat{\rho}_m\}$, the ML estimates. So ML estimates may be found from a nonlinear regression algorithm that seeks the zeros of $\mu(A_m,\rho_m)$, m=1,2,...,p.

Linear Time Invariant (Markovian) Representation: Let $\{y_t\}$ be the ARMA (p,p) sequence of Eq. (2). The standard form or Markovian state space representation for $\{y_t\}$ is

$$y_t = \hat{y}_{t/t-1} + u_t \qquad \underline{c}' = [1 \;\; 0 \ldots 0]$$

$$\hat{y}_{t/t-1} = \underline{c}' x_t$$

$$x_t = A x_{t-1} + \underline{h}_1 u_{t-1} \qquad A = \begin{bmatrix} 0 & 1 & & 0 \\ & & \ddots & \\ & & & 1 \\ -a_p & & \cdots & -a_1 \end{bmatrix} \quad \underline{h}_1 = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}$$

$$x_0 : N(0, Q_0)$$

u_t : i.i.d. $N(0,\sigma^2)$ r.v.s.

$$x_t = \begin{bmatrix} \hat{y}_{t/t-1} \\ y_{y+1/t-1} \\ \vdots \\ y_{t+p-1/t-1} \end{bmatrix} \qquad \begin{bmatrix} 1 & & & \\ a_1 & 1 & & \\ a_2 & a_1 & 1 & \\ \vdots & & & \\ a_p & & & 1 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_p \end{bmatrix} = \begin{bmatrix} 1 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

In this representation it is important to note that the p-dimensional state x_t is a vector of s-step predictors $\hat{y}_{t+s/t-1}$ (s=0,1,...,p-1) based on the infinite past $(y_{t-1}, y_{t-2}, \ldots)$:

$$\hat{y}_{t+s/t-1} = E[y_{t+s} / y_{t-1}, y_{t-2}, \ldots]$$

The unit pulse response sequence $\{h_t\}$ and correlation sequence $\{r_t\}$ provide invaluable first- and second-order descriptors for the ARMA process:

$$h_t = \begin{cases} 0 & , t<0 \\ 1 & , t=0 \\ \underline{c}'A^{t-1}\underline{h}_1 & , t\geq 1 \end{cases}$$

$$r_t = \underline{c}'A^t Q_0 \underline{c} + \sigma^2 h_t \ , \ t \geq 0$$

$$r_{-t} = r_t$$

$$Q_0 = AQ_0A' + \sigma^2 \underline{h}_1\underline{h}_1'$$

Figure 1 shows feedback diagrams for this process model and the corresponding predictor structure. Note that both diagrams are time-invariant.

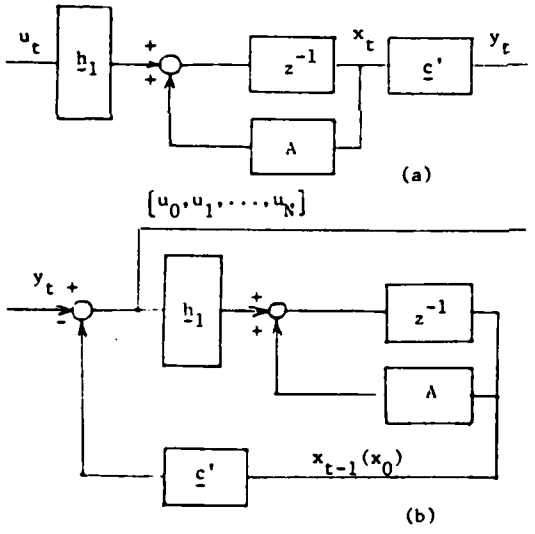


Fig. 1. Markovian State Space Model for ARMA (p,p) Process: (a) model, (b) predictor

Linear Time-Varying (Innovations) Representation: The time-varying, or innovations, representation for $\{y_t\}_0^\infty$ is

$$y_t = y_{t/t-1} + u_t$$

$$\hat{y}_{t/t-1} = \underline{c}'x_t$$

$$x_t = Ax_{t-1} + \underline{k}_{t-1}u_{t-1}$$

$$x_0 = 0$$

$$\underline{c}' = [1\ 0\ \ldots\ 0]$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ & & & \ddots & 1 \\ -a_p & & \ldots & & -a_1 \end{bmatrix} \quad \underline{k}_t = \begin{bmatrix} k_t^{(1)} \\ \vdots \\ k_t^{(p)} \end{bmatrix}$$

u_t: indep. $N(0, v_t)$ r.v.s.

$$x_t = \begin{bmatrix} \hat{y}_{t/t-1} \\ \hat{y}_{t+1/t-1} \\ \vdots \\ \hat{y}_{t+p-1/t-1} \end{bmatrix}$$

$$\underline{k}_t v_t = -AP_t\underline{c} + AQ_0\underline{c} + \sigma^2\underline{h}_1$$

$$v_t = r_0 - \underline{c}'P_t\underline{c}$$

$$P_t = E(x_t x_t')$$

In this representation the p-dimensional state x_t is a vector of s-step predictors $\hat{y}_{t+s/t-1}$ $(s=0,1,\ldots,p-1)$ based on the finite past $\{y_{t-1}, y_{t-2}, \ldots, y_0\}$:

$$\hat{y}_{t+s/t-1} = E[y_{t+s}/y_{t-1}, y_{t-2}, \ldots, y_0]$$

The time-varying unit pulse response sequence $\{h_i^t\}$ and the (generally non-stationary) correlation sequence $\{r_i^s\}$ provide first- and second-order descriptors for the ARMA process:

$$h_i^t = 0 \quad i < 0$$

$$h_0^t = 1$$

$$h_i^t = \underline{c}'A^{i-1}k_t \quad i > 0$$

$$r_i^t = \underline{c}'A^i P_t c + c'A^{i-1}\underline{k}_t v_t$$

$$P_{t+1} = AP_tA' + \underline{k}_t\underline{k}_t' v_t$$

$$r_i^t = E[y_t\, y_{t+i}]$$

With $\underline{k}_t v_t$ chosen to be

$$\underline{k}_t = -AP_t\underline{c} + AQ_0\underline{c} + \sigma^2\underline{h}_1 \qquad (6)$$

then $r_i^t = r_i$ and $\{y_t\}_0^\infty$ is wide-sense stationary with the same correlation function obtained in the previous section. Figure 2 shows feedback diagrams for this model and the corresponding predictor structure. Note that both diagrams are time-varying. The innovations model is simply a time-varying model that starts from zero initial conditions and, nevertheless, generates a stationary sequence.

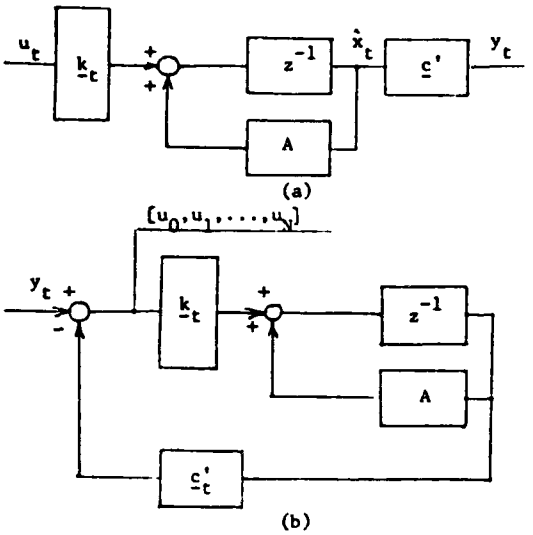


Fig. 2. Innovations Representation for ARMA (p,p) Process: (a) model, (b) predictor

Exact Likelihood in the Markovian Representation: From the predictor structure of Figure 1 we note that, given x_0, Y is distributed as follows:

$$f(Y/x_0) = \prod_{t=0}^{N} N_{y_t}(\hat{y}_{t/t-1}, \sigma^2)$$

Here N_{y_t} is the normal density

$$N_{y_t}(\hat{y}_{t/t-1}, \sigma^2) = (2\pi\sigma^2)^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma^2}(y_t - \hat{y}_{t/t-1})^2\right)$$

Caution: $\hat{y}_{t/t-1}$ is a function of x_0! But x_0 is distributed as follows:

$$x_0 : N_{x_0}(0, Q_0) = (2\pi)^{-p/2} |Q_0|^{-1/2} \cdot \exp\left(-\frac{1}{2} x_0' Q_0^{-1} x_0\right)$$

So the unconditional density of Y is

$$f(Y) = \int dx_0 N_{x_0}(0, Q_0) \prod_{t=0}^{N} N_{y_t}(\hat{y}_{t/t-1}, \sigma^2) \quad (7)$$

The exact likelihood is $\ell_N(Y) = \ln f(Y)$. In Section 5 we show how this representation may be used to obtain a variety of approximate likelihood functions.

Exact Likelihood in the Innovations Representation: From the predictor structure of Figure 2 we see that Y is distributed as follows:

$$f(Y) = \prod_{t=0}^{N} N_{y_t}(\hat{y}_{t/t-1}, v_t) \quad (8)$$

where $\hat{y}_{t/t-1}$ now depends only on the finite past and v_t is a time varying prediction error variance. The exact likelihood is $\ell_N(Y) = \ln f(Y)$.

Connections: A Chapman Kolmogorov Equation and Initial Conditions: Comparing Eqs. (7) and (8) it is clear that the innovations representation has solved a very important Chapman-Kolmogorov equation:

$$\int dx_0 \, N_{x_0}(0, Q_0) \prod_{t=0}^{N} N_{y_t}(\hat{y}_{t/t-1}(x_0), \sigma^2) = \prod_{t=0}^{N} N_{y_t}(\hat{y}_{t/t-1}, v_t) \quad (9)$$

(We have used the notation $\hat{y}_{t/t-1}(x_0)$ on the LHS to distinguish between the two $\hat{y}_{t/t-1}$.) This is one interpretation. But note f(Y) may be written

$$f(Y) = f(Y/x_0) f(x_0)/f(x_0/Y)$$

What this means is that solution of the Chapman-Kolmogorov equation is tied up in the solution for $f(x_0/Y)$, the *a posteriori* density for the initial condition x_0, given observations Y. This is fundamental. Always the problem is initial conditions. In Section 5 we derive a closed form expression for $f(x_0/Y)$, and thus derive an exact time-invariant realization of the likelihood function.

The Innovations Representation is Equivalent to a Sequence of AR(t) Models of Increasing Orders: The innovations representation may be interpreted as a sequence of AR(t) filters in disguise. This interpretation provides for another constructive approach to recursive computation of exact likelihood. The key is to use the properties of the Toeplitz correlation matrix R_N to achieve a sequential computation in terms of dimension.

Recall the original likelihood expression

$$\ell_N(Y) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|R_N| - \frac{1}{2}Y'R_N^{-1}Y$$

Partition R_N to obtain an equivalent partition of the inverse,

$$R_N = \begin{bmatrix} r_0 & r' \\ r & R_{N-1} \end{bmatrix} \qquad R_N^{-1} = \begin{bmatrix} 1/\alpha_N & a'N/\alpha_N \\ a_N/\alpha_N & [R_{N-1} + \frac{rr'}{r_0}]^{-1} \end{bmatrix} \quad (10)$$

with $\alpha_N = r_0 - r'R_{N-1}^{-1}r = |R_N|/|R_{N-1}|$. The vector a_N is easily recognized as the set of optimal parameters for an autoregressive model AR(N) of order N fitted on the given model (ARMA) correlation R_N. As a consequence the likelihood is decomposed according to (11):

$$\ell_N = \ell_{N-1} - \frac{1}{2}\log\alpha_N - \frac{1}{2}\frac{\hat{u}_N^2}{\alpha_N} - \frac{1}{2}\log 2\pi \quad (11)$$

Here $\hat{u}_N$ is the prediction error of the AR(N) model measured on the data and α_N the corresponding variance.

This approach is obviously equivalent to the preceding innovation representation and is merely another way of computing the innovations sequence. The whitening condition is here replaced by the orthogonality condition satisfied by the optimal AR estimated model. In the following, one of the two approaches will more naturally give rise to a time sequential computation of likelihood and the other to a batch implementation.

3. RECURSIVE COMPUTATION OF LIKELIHOOD FOR BATCH DATA

In many applications the measurements $y_0 \ldots y_N$ are readily available, and can be processed as a batch. Moreover, the likelihood computation for a given model is generally imbedded in a larger iterative procedure for optimizing the model (such as nonlinear programming) which is not fitted for real-time implementation. An efficient batch method for evaluating the likelihood is therefore often desirable.

A first solution is given by the expression for R_N^{-1} using the autoregressive estimates of successive orders. The appropriate Cholesky factorization is

$$R_N^{-1} = \begin{bmatrix} 1 & & & \\ a_1^N & 1 & & \\ \vdots & & \ddots & \\ a_N^N & \cdots & a_1^1 & 1 \end{bmatrix} \begin{bmatrix} 1/\alpha_N & & & \\ & \ddots & & \\ & & & 1/\alpha_0 \end{bmatrix} \begin{bmatrix} 1 & a_1^N & \cdots & a_N^N \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \quad (12)$$

where a_i^j represent the coefficients of the optimal AR prediction of order j and α_j the corresponding variance. The sum of squares in the likelihood is then obtained as the squares of the successive outputs of a time varying, order increasing MA filter fed with the data to generate the prediction residuals:

$$z_t^t = \sum_{i=0}^{t} a_i^t \, y_{t-i} \quad (13)$$

This computation is started with zero initial conditions and conveniently weighted by the inverse variance α_t^{-1}.

A seemingly more convenient implementation, first proposed by Kailath et al. [15], makes use of a now classical expression of the inverse of a Toeplitz matrice [16] in terms of the coefficients of the order N predictor (i.e. first column of the triangular factorization in (12)),

$$R_N^{-1} = \begin{bmatrix} a_0 & & & \\ a_1 & a_0 & & \\ \vdots & & \ddots & \\ a_N & \cdots & a_1 & a_0 \end{bmatrix} \begin{bmatrix} a_0 & a_1 & \cdots & a_N \\ & a_0 & & \\ & & \ddots & \\ & & & a_0 \end{bmatrix} - \begin{bmatrix} 0 & & & \\ a_N & 0 & & \\ & & \ddots & \\ a_1 & a_N & \cdots & 0 \end{bmatrix} \begin{bmatrix} 0 & & & \\ a_1 & 0 & & \\ & & \ddots & \\ a_N & a_{N-1} & \cdots & 0 \end{bmatrix} \quad (14)$$

where $a_i = a_i^N \cdot \sqrt{\alpha_N}$

This representation is suitable for time invariant implementation using two reciprocal MA filters of given order N and N-1:

$$z_t^N = \sum_{i=0}^{t} a_i \, y_{t-i}$$

$$z_t^{*N} = \sum_{i=1}^{t} a_{N-i+1} \, y_{t-i} \quad (15)$$

The outputs are then squared, subtracted, weighted and accumulated to compute the desired likelihood. The major limitation of this method is evidently the restriction to a fixed sample size N.

But the opposition between the two implementations of Eqs. (13) and (15) is only apparent. The time-variation in (13) is more meaningfully interpreted as an order adjustment at each sample. The two methods can be imbedded in the same lattice structure using only the reflection coefficients $\gamma_j = a_j^j$ (i.e. last row of the triangular factorization in (12))[17]. The preceding quantities z_t^j are then available at the output of the various sections of the lattice according to Figure 3.

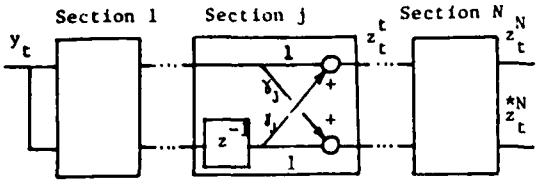


Fig. 3. Use of an AR lattice for computing exact likelihood for ARMA process

This structure is well known for its excellent sensitivity properties and allows at the same time a convenient sequentialization of the computation by adding new sections to the fixed preceding ones.

The computational requirement for the above method is the fitting of an order N AR predictor to the correlation sequence of the given ARMA model under test. This is conveniently achieved using the classical Levinson algorithm in order of N^2 operations instead of N^3. But still N is large in many applications (e.g. ~ 256) and the correlation must be carefully computed (even in this noise free case) to ensure the necessary stability of the predictions.

4. RECURSIVE COMPUTATION OF LIKELIHOOD FOR REAL-TIME DATA

The state space formulation gives a nice conceptual way of generating the innovation sequence in the well understood framework of Kalman filtering. The formulation becomes particularly interesting when we note that there exist efficient algorithms for computing the Kalman gains. This efficiency is not achieved using standard algorithms that ignore the fine structure of the data. Here the data are output correlation coefficients for an ARMA process, and the state space equations describing them must have an input/output counterpart.

More precisely the convolutional form for the time-varying innovation representation is

$$y_t = \sum_{i=1}^{\infty} h_i^t \, u_{t-i} + h_0 \, u_t$$

which relates to the state space description by

$$x_t = \sum_{i=1}^{\infty} \underline{h}_i^t \, u_{t-i} \;,\; \underline{h}_i^t = A^{i-1} \underline{k}_t \;,\; h_i^t = c' A^{i-1} \underline{k}_t$$

This impulse response description will now be used.

Instead of using the Cholesky factorization of the inverse as in (12) the correlation matrix itself may be uniquely decomposed into an upper and a lower (transposed) factor.

$$R_t = \begin{bmatrix} s_0^t & s_p^{t-p} & \cdots & s_0^t \\ & s_0^{t-p} & \cdots & s_{t-p}^0 \\ & & & \\ & & s_0^1 & s_1^0 \\ & & & s_0^0 \end{bmatrix}^2 \qquad (16)$$

Here the notation "2" stands for the product by its transpose. The quantities s_i^j are readily interpreted as cross correlations between the y_t and their prediction error for a predictor of order j. As the r_i for i>q satisfy the AR recursion with coefficients a_i, it is also the case for the s_i^j i>q. After the normalisation by $1/s_0^j$, the s_i^j are the coefficients of a time varying impulse response. By taking the first (p+1) rows and columns of R_t, it follows that:

$$\begin{bmatrix} r_0 & \cdots & r_p \\ \vdots & & \vdots \\ r_p & \cdots & r_0 \end{bmatrix} = \begin{bmatrix} s_0^t & \cdots & s_p^{t-p} & \cdots & s_t^0 \\ 0 & \ddots & & & \vdots \\ & & s_0^{t-p} & \cdots & s_{t-p}^0 \\ 0 & 0 & & & \end{bmatrix}^2 \qquad (17)$$

As a consequence the time-varying system with impulse response h_i^t fed with white noise of variance v_t both defined such that

$$h_i^t = s_i^t / s_0^t \quad \text{and} \quad v_t = s_0^t \qquad (18)$$

reproduces the stationary correlation of the process. The h_i^t for i=1,p coincide with the components of the Kalman gain k_t and v_t is the corresponding variance of the innovations.

An efficient algorithm has been provided for factoring R_t according to (16) [18]. Moreover, it has been proven to generate a minimum phase model [19] and can be implemented using fixed point arithmetic [20]:

$$\begin{aligned} \gamma_t &= -s_{-t-1}^t / s_0^t \\ s_0^{t+1} &= (1-\gamma_t^2) s_0^t \\ s_i^{t+1} &= s_i^t + \gamma_t s_{-t-i-1}^t \quad \forall\, i \end{aligned} \qquad (19)$$

In this particular case, only the s_i^{t+1} for i=0,p are useful. They are computed using p nonzero values for i<0, so that the required number of multiplications per step is approximately 2p. The values for i>p may be deduced if necessary from the recursion with coefficients a_i.

Moreover it may be noted that the most important

role in the behavior of the model is played by the MA part. For an AR-only process, the Kalman gain would converge in p steps (p first AR predictors) but for an MA process the gain converges at infinity. It seems therefore interesting to consider the AR and MA parts separately instead of constructing the ARMA correlation sequence. This can be done by extending the AR part of the equation in order to include the zero initial conditions and deduce the time varying MA correlation matrix V for the given stationary ARMA output correlation R_N according to (20):

$$\begin{bmatrix} a_0 & a_1 \cdots a_p & & \\ & a_0 & & \\ & & \ddots & \\ & & a_0 \cdots a_p & \\ & & & \ddots \\ & & & a_0 \end{bmatrix} R_N \begin{bmatrix} a_0 & & & \\ \vdots & \ddots & & \\ a_p & & a_0 & \\ & \ddots & \vdots & \ddots \\ & & a_p \cdots a_0 & \end{bmatrix} = V \tag{20}$$

Due to the zero initial conditions, the V matrix is no longer Toeplitz but is still a banded matrix and can be computed using a_i and b_i. A generalization of Bauer's Algorithm [21] may be applied in this time-varying case [22] to factor V and retrieve the Kalman gains. The algorithm then requires approximately $q^2/2$ multiplies per step and can be interesting for small q. A closer examination of V, shows that it is Toeplitz except for a pxp right bottom matrix and it is the feeling of these authors that the computations could be reduced using other choices for the initial AR and MA parts.

5. INITIAL CONDITIONS

We return to the ARMA difference equation:

$$y_t + a_1 y_{t-1} + \ldots + a_p y_{t-p} = u_t + b_1 u_{t-1} + \ldots + b_p u_{t-p}$$

Let us establish the following data conventions:

$$Y_{-p} = \begin{bmatrix} y_{-p} \\ \vdots \\ y_{-1} \end{bmatrix} \quad Y_0 = \begin{bmatrix} y_0 \\ \vdots \\ y_{p-1} \end{bmatrix} \quad U_{-p} = \begin{bmatrix} u_{-p} \\ \vdots \\ u_{-1} \end{bmatrix} \quad U_0 = \begin{bmatrix} u_0 \\ \vdots \\ u_{p-1} \end{bmatrix}$$

And

$$A = \begin{bmatrix} a_0 & & \\ a_1 & a_0 & \\ \vdots & \ddots & \\ a_{p-1} & \cdots & a_0 \end{bmatrix} \quad H_p = \begin{bmatrix} h_0 & & \\ h_1 & h_0 & \\ \vdots & \ddots & \\ h_{p-1} & \cdots & h_0 \end{bmatrix} \quad B = \begin{bmatrix} b_0 & & \\ b_1 & b_0 & \\ \vdots & \ddots & \\ b_{p-1} & \cdots & b_0 \end{bmatrix}$$

$$A^* = \begin{bmatrix} a_p & \cdots & a_1 \\ & \ddots & \vdots \\ & & a_p \end{bmatrix} \quad B^* = \begin{bmatrix} b_p & \cdots & b_1 \\ & \ddots & \vdots \\ & & b_p \end{bmatrix}$$

The (pxp) Toeplitz matrix of unit pulse responses satisfies the linear equation

$$A H_p = B \tag{21}$$

Special cases are

MA : $H_p = B$ AR : $A H_p = I$

We may summarize initial conditions by establishing the following linear dependence:

$$A^* Y_{-p} + A Y_0 = U_{-p} + B U_0 \tag{22}$$

The special cases are

MA: AR:

$$Y_0 = B^* U_{-p} + B \qquad A^* Y_{-p} + A Y_0 = U_0$$

From the time-invariant state space representation, we have initial conditions:

$$x_0 = Y_0 - H_p U_0 \qquad Q_0 = R_p - H_p H_p' \tag{23}$$

The special cases are

MA: AR:

$$Y_0 = x_0 + B U_0 \qquad Y_0 = x_0 + A^{-1} U_0$$

$$x_0 = B^* U_{-p} \qquad x_0 = -H_p A^* Y_{-p}$$

5.1 Approximate Likelihood: Fixed Initial Conditions:

Recall the likelihood function of Eq. (7). With the normal density $N_{x_0}(0,Q_0)$ replaced by the dirac density $\delta(x_0-\hat{x}_0)$, we have the approximate likelihood:

$$f(Y/\hat{x}_0) = \prod_{t=0}^{N} N_{y_t}(y_{t/t-1}(\hat{x}_0), \sigma^2)$$

Question: How to choose $\hat{x}_0$?

Method 1: With $\hat{x}_0 = Y_0$ we are assuming $U_0 = 0$ in (23). This implies $Q_0 = R_p$. The approximate likelihood is simply obtained by summing squared residuals in Figure 1, with $\hat{x}_0$ fixed at Y_0. The special cases are

MA: AR:

$$Y_0 = B^* U_{-p} \qquad A Y_0 = A^* Y_{-p}$$

In each of these special cases, initial conditions are manufactured outside the data interval with a backward predictor. In the AR case Y_{-p} is a backward prediction from Y_0. The resulting approximate likelihood function reproduces the equations of the covariance method of linear prediction.

Method 2. With $\hat{x}_0 \equiv 0$ we have $Y_0 = HU_0$. This implies $Q_0 \equiv 0$. Again the approximate likelihood is obtained by summing residuals in Figure 1 with $\hat{x}_0$ fixed at 0. The special cases are

MA:	AR:
$U_{-p} = 0$	$Y_{-p} = 0$

In this method initial conditions are set to zero outside the data interval. In the AR case the resulting equations correspond to the prewindowed method of linear prediction.

5.2 Approximate Likelihood: MAP Estimation of Initial Conditions:

As a second alternative to approximating exact likelihood, consider maximizing the joint density of the data Y and the initial conditions x_0:

$$\max_{x_0} f(Y,x_0)$$

Write out the time-invariant (Markovian) state-space equation for t=0,1,...,N:

$$Y = \mathcal{O}x_0 + H_N U$$

$$Y' = [y_0 \cdots y_N] \qquad U' = [u_0 \cdots u_N]$$

$$\mathcal{O}' = [C\ ,\ A'C\ ,\ \ldots,\ (A')^N C]$$

We have the following linear dependence

$$U = L\,Y - Mx_0\ ,$$

with

$$L = H_N^{-1} \quad \text{and} \quad M = H_N^{-1}\,\mathcal{O}$$

The matrix $\mathcal{O}$ is the observability matrix.

The joint density of Y and x_0 is obtained from the density of U:

$$f(Y,x_0) = (2\pi\sigma^2)^{-(N+1)/2}\exp\{-\frac{1}{2\sigma^2}\,g(Y,x_0)\}$$

$$g(Y,x_0) = (LY-Mx_0)'(LY-Mx_0)$$

Note $g(Y,x_0)$ may be written

$$g(Y,x_0) = g(Y,\hat{x}_0) + (x_0-\hat{x}_0)'M'M(x_0-\hat{x}_0)$$

with

$$M'M\hat{x}_0 = M'LY$$

So $f(Y,\hat{x}_0)$ is

$$f(Y,\hat{x}_0) = (2\pi\sigma^2)^{-(N+1)/2}\exp\{-\frac{1}{2\sigma^2}\,(LY-M\hat{x}_0)'(LY-M\hat{x}_0)\} \qquad (24)$$

The corresponding likelihood is $\ell(Y,\hat{x}_0) = \ell nf(Y,\hat{x}_0)$.

The AR case specializes nicely:

$$\hat{x}_0 = Y_0 - (H_p H_p')R_p^{-1}\,Y_0$$

Equivalently,

$$R_p^{-1}\,Y_0 = Q_0^{-1}\,\hat{x}_0$$

or

$$A^*Y_{-p} + H_p^{-1}\,Q_0\,R_p^{-1}\,Y_0 = 0$$

Note this reduces to the covariance method of linear prediction under the approximation $Q_0R_p^{-1} = I$. The corresponding likelihood is

$$\ell(Y,\hat{x}_0) = \frac{1}{2\sigma^2}\sum_{t=p}^{N}\Big(y_t - \sum_{i=1}^{p} a_i y_{t-i}\Big)^2 + \frac{1}{2}Y_0'R_p^{-1}Y_0 \qquad (25)$$

The second term on the RHS represents a correction to the usual covariance method of linear prediction.

5.3 Exact Likelihood:

Write the joint density of the data Y and the initial conditions x_0 as

$$f(Y,x_0) = f(Y/x_0)f(x_0) = f(x_0/Y)f(Y)\ .$$

Here

$f(x_0)$: prior density on x_0 : $N_{x_0}(0,Q_0)$

$f(Y/x_0)$: conditional density :

$$\prod_{t=0}^{N} N_{y_t}(\hat{y}_{t/t-1}(x_0),\sigma^2)$$

$f(x_0/Y)$: *a posteriori* density for x_0, given data Y

$f(Y)$: unconditional density of Y

Return to the joint density $f(Y,x_0)$:

$$f(Y,x_0) = (2\pi\sigma^2)^{-(N+2)/2}\exp\{-\frac{1}{2\sigma^2}\, g(Y,\hat{x}_0) - \frac{1}{2\sigma^2}\,(x_0-\hat{x}_0)'M'M(x_0-\hat{x}_0)\}$$

It follows from a factorization theorem that

$$f(x_0/Y) = (2\pi\sigma^2)^{-p/2}|M'M|^{\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}\,(x_0-\hat{x}_0)'M'M(x_0-\hat{x}_0)\}$$

Now write

$$f(Y) = f(Y/x_0)f(x_0)/f(x_0/Y)$$

$$\ell(Y) = \ell(Y/x_0) + \ell(x_0) - \ell(x_0/Y)$$

$$\ell(Y) = -\frac{1}{2\sigma^2}\sum_{t=0}^{N}(y_t-\hat{y}_{t/t-1}(x_0))^2 - \frac{1}{2}\,x_0\,Q_0x_0' + \frac{1}{2}\log|M'M| - \frac{1}{2\sigma^2}\,(x_0-\hat{x}_0)'M'M(x_0-\hat{x}_0) \qquad (26)$$

The first summation may be obtained from the time-invariant predictor in Figure 1, starting from any x_0. But the choice must be fixed up with the remaining terms. Several choices are of interest : $x_0 = \hat{x}_0$, $x_0 = 0$.

6. CONCLUSIONS

Maximum likelihood (ML) is an attractive parametric method for fitting models to observed data. When the data is a time series record and the underlying model is ARMA, then ML becomes a parametric method for identifying linear models. In the parlance of signal and system theory, ML becomes a method of identifying rational discrete-time systems from output data only.

In general ML leads to nonlinear equations in the parameters to be identified. This is reflected in all of our expressions for exact likelihood. By making special assumptions about initial conditions one can obtain in the AR case quadratic approximations to exact ML that lead to such methods as the pre-windowed and covariance methods of linear prediction.

If no approximations are to be made then two problems arise: (1) efficient computation of exact likelihood for each set of candidate ARMA parameters, and (2) evaluation of gradients and/or Hessians for efficient iteration to a solution.

In this paper we have focused on problem (1). Several approaches to the computation of exact likelihood have been analyzed and compared. As in previous contributions to this problem, a central role is played by the innovation of the ARMA process. Other derivations of exact likelihood have proceeded within the framework of Kalman filtering where the innovations play an essential part. But innovations need not be tied to Kalman filtering. In this paper we have emphasized the innovations representation for the ARMA process itself. This leads to a straightforward computation of Kalman gains designed to generate a stationary correlation sequence in spite of the zero-initial conditions imposed on the model. The state-space consists of one-step through p-step predictions based on the finite past. An interesting interpretation is that the time-varying Kalman gains comprise the time-varying MA part of an ARMA model with fixed AR part.

The innovations of a stationary ARMA process need not be interpreted in the context of state space models. They may also be interpreted as a sequence of independent random variables obtained by sequentially whitening the correlated data. The whitening transformation is derived by triangularizing an inverse covariance matrix. The triangularization procedure is equivalent to fitting a sequence of AR models to ARMA data. Stable and efficient algorithms, which compete favorably with fast Kalman algorithms, are available.

The problem of computing exact likelihood may also be cast in such a way that initial conditions on a time-invariant predictor play an important role. The initial conditions may be pinned at an arbitrary value provided subsequent corrections are applied. This formulation shows the way to approximations of exact likelihood that may be superior to existing ones. The same approach could have been applied to final conditions, providing a completely symmetrical setting in which to discuss forward and backward prediction outside the data window.

As mentioned previously, this paper is focused on one aspect of ML identification of ARMA models: computation of exact likelihood. It is our feeling that some progress is still to be made on this problem to reduce complexity. But just as important for overall maximization of likelihood would be efficient procedures for evaluating gradients and Hessians in the framework of time-varying innovations representations or time-invariant predictor representations.

REFERENCES

[1] H. Akaike, "Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes," Ann. Inst. Statist. Math., 26, p. 363 (1974).

[2] P. Newbold, "The exact likelihood function for a mixed autoregressive-moving average process," Biometrika, 61, 3, p. 423 (1974).

[3] D. Osborne, "Maximum likelihood estimation of moving average processes," Ann. Econ. & Soc. Meas. 5, p. 75 (1976).

[4] A. C. Harvey and G.D.A. Phillips, "Maximum likelihood estimation of regression models with autoregressive-moving average disturbances," Biometrika, 66, 1, p. 49 (1979).

[5] C. F. Ansley, "An algorithm for the exact maximum likelihood of a mixed autoregressive-moving average process," Biometrika, 66, p. 59 (1979).

[6] J. G. Pearlman, "An algorithm for the exact likelihood of a high-order autoregressive-moving average process," Biometrika, 67, 1, p. 232 (1980).

[7] R. H. Jones, "Maximum likelihood fitting of ARMA models to time series with missing observations," Technometrics, to appear 1980.

[8] G.E.P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco (1976).

[9] F. C. Schweppe, "Evaluation of likelihood functions for Gaussian signals," IEEE Trans. Info. Theory, IT-11, 1, p. 61 (1965).

[10] M. Morf, G. S. Sidhu, and T. Kailath, "Some new algorithms for recursive estimation on constant linear discrete-time systems," IEEE Trans. Auto. Contr., AC-19, p. 315 (1974).

[11] J. Perl and L. L. Scharf, "Covariance-Invariant Digital Filtering," IEEE Trans. Acoust., Speech, Signal Proc., ASSP-25, 2, p. 143 (1977).

[12] T. W. Anderson, "Statistical inference for covariance matrices with linear structure," Multivariable Analysis II (P. R. Krishnaiah, ed.).

[13] P. R. Aigrain and E. M. Williams, "Synthesis of n-reactive networks for desired transient response," J. Appl. Physics, 20, p. 297 (1949).

[14] L. L. Scharf and L. W. Nolte, "Likelihood ratios for sequential hypothesis testing on Markov sequences," IEEE Trans. Inform. Theory, IT-23, 1, p. 101 (Jan. 1977).

[15] T. Kailath, B. Levy, L. Ljung, M. Morf, "Fast time-invariant implementation of Gaussian signal detectors," IEEE Trans. on IT, vol. IT24, no. 4, July 1978.

[16] M. Siddiqui, "On the inversion of the sample covariance matrix in the stationary AR process," Ann. math. stat., no. 28, pp. 585-588 (1958).

[17] J. Makhoul, "A class of all zero lattice digital filters, properties and applications," IEEE Trans. on ASSP, vol. 26, no. 4, pp. 304-314 (August 1978).

[18] J. Rissanen, "Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with application to factoring positive matrices," Mathematics of Computation, vol. 23, no. 121, pp. 147-154 (Jan. 1973).

[19] J. LeRoux, Y. Grenier, "An iterative procedure for moving average model estimation," IEEE Int. Cont. on ASSP, pp. 614-617 (April 1980).

[20] J. LeRoux, C. Gueguen, "A fixed point computation of the partial correlation coefficients," IEEE Trans. on ASSP, vol. 25, no. 3, pp. 257-259 (June 1977).

[21] F. L. Bauer, "Iterstionsverphren jur Hurwitz-zerlegang eines polynoms," Archiv der Elektrischen Ubertragugn, band 9, heft 6, pp. 285-290 (June 1955).

[22] J. Rissanen, L. Barbosa, "Properties of infinite covariance matrices and stability of optimal predictors," Information Science, no. 1, pp. 221-236 (1969).

PARAMETRIC SPECTRUM MODELLING: A SIGNAL PROCESSING PERSPECTIVE

Louis L. Scharf
Claude J. Gueguen
Jean-Pierre Dugre

Presented at:

1st IEEE Workshop on Spectrum Analysis
McMaster University, Hamilton, Ontario
August 1981

PARAMETRIC SPECTRUM MODELLING: A SIGNAL PROCESSING PERSPECTIVE

Louis L. Scharf, Sr. Member IEEE
Electrical Engineering Department, Colorado State University
Fort Collins, Colorado 80523

Claude J. Gueguen, Sr. Member IEEE
Ecole Nationale Superieure de Telecommunications
46 rue Barrault 75634 Paris

Jean-Pierre Dugré
Electrical Engineering Department, Colorado State University
Fort Collins, Colorado 80523

Abstract

In this paper we present a general framework for deriving and interpreting analysis and synthesis spectra of the autoregressive (AR) and moving average (MA) type. By analyzing MA and AR linear transformations of finite-dimensional data records we derive MA type spectra that are direct analogs of the AR type spectra associated with the maximum likelihood method (MLM) and the maximum entropy method (MEM) of spectrum analysis. Asymptotically the MA theory is tied up with Wold's decomposition in the same way the AR theory is tied up with Kolmogorov's whitening theory.

By parametrizing the MA and AR type spectra we obtain a variety of spectrum models that trade off resolution and power fidelity. We propose J-divergence as an attractive order fitting rule and show how it relates to final prediction error (FPE) and Akaike's information criterion (AIC).

1.0 Introduction

The concept of wide-sense stationarity seems to underly the very notion of a spectrum. It is a mistake however, to conclude that one should identify only stationary models when estimating spectra. The problem with stationary models is that initial conditions manifest themselves as nuisance parameters that must either be estimated or averaged over. In nonstationary models the initial conditions manifest themselves as time variations (as in innovations representations) or as order-increases (as in lattice representations). In either case the initial conditions may be absorbed naturally into the theory.

In this paper we present a general framework for deriving and interpreting analysis and synthesis spectra of the autoregressive (AR) and moving average (MA) type. By analyzing MA and AR linear transformations of finite-dimensional data records we derive MA type spectra that are direct analogs of the AR type spectra associated with the maximum likelihood method (MLM) and the maximum entropy method (MEM) of spectrum analysis. Asymptotically the MA theory is tied up with Wold's decomposition in the same way the AR theory is tied up with Kolmogorov's whitening theory.

By parametrizing the MA and AR type spectra we obtain a variety of spectrum models that trade off resolution and power fidelity. We propose J-divergence as an attractive order fitting rule and show how it relates to final prediction error (FPE) and Akaike's information criterion (AIC).

For an autoregressive moving average (ARMA) time series, a fast impulse response algorithm may be used in conjunction with an innovations representation or a synthesis lattice to efficiently realize an MA type synthesis of ARMA data. Conversely a fast Levinson/Durbin algorithm may be used in conjunction with a Kalman predictor or an analysis lattice to efficiently realize an AR type analysis of ARMA data. In the concluding section we outline how these observations lead directly to a derivation of fast and exact likelihood for ARMA time series.

This work supported in part by the Office of Naval Research, Statistics and Probability Branch, Arlington, VA, under Contract N00014-75-C-0518 and the Army Research Office, Research Triangle Park, NC, under Contract DAAG29-79-C-0176.

2.0 Linear Filters and Representation of Stationary Sequences

Let $\{y_t\}$ denote a real, zero-mean, wide-sense stationary sequence with real ℓ_2 correlation sequence $\{r_t\}$. This sequence has factorization

$$r_t = \sum_{n=0}^{\infty} h_n h_{n+|t|} \quad , \quad \forall\, t$$

We call $\{h_n\}$ the impulse response sequence, for reasons to become clear.

Whenever we speak of $\{y_t\}$ we have in mind the triple $(\{y_t\},\{h_t\},\{r_t\})$.

2.1 MA(∞) or Wold Decomposition

The Wold decomposition for the sequence $\{y_t\}$ is the infinite moving average (MA(∞))

$$y_t = \sum_{n=0}^{\infty} h_n u_{t-n}$$

u_t : white sequence (zero-mean, unit variance r.v.s.)

The coefficients $\{h_n\}$ are called the MA(∞) filter coefficients.

First-Order Descriptors. Replace $\{u_t\}$ by the impulse sequence $\{\delta_t\}$ to obtain the impulse response

$$h_t = \begin{cases} h_t & , \quad t \geq 0 \\ 0 & , \quad \text{ow} \end{cases}$$

The complex frequency response associated with this impulse response is[1]

$$h(\omega) = \lim_{t\to\infty} \underline{h}^H C_t(\omega)$$

$$C_t^H(\omega) = (1, e^{+j\omega}, \ldots, e^{+j(t-1)\omega})$$

$$\underline{h}^H = (h_0, h_1, \ldots, h_{t-1})$$

[1] Superscript H denotes Hermitian transpose.

Second-Order Descriptors. The correlation sequence is

$$r_t = \sum_{n=0}^{\infty} h_n h_{n+|t|}$$

The corresponding power spectral density (or magnitude-squared frequency response) is

$$r(\omega) = |h(\omega)|^2$$

MA(q) Case. In the MA(q) case the moving average representation terminates and we may write

$$y_t = \sum_{n=0}^{q} h_n u_{t-n}$$

The resulting specializations in the first- and second-order descriptors are obvious.

2.2 AR(∞) or Kolmogorov Representation

The Komogorov representation for $\{y_t\}$ is the infinite autoregression (AR(∞)):

$$\sum_{n=0}^{\infty} a_n y_{t-n} = u_t \quad ; \quad a_0 \neq 0$$

The coefficients $\{a_n\}$ are called the AR(∞) filter coefficients.

First-Order Descriptors. Replace $\{u_t\}$ by the impulse sequence $\{\delta_t\}$ to obtain the impulse response

$$\sum_{n=0}^{\infty} a_n h_{t-n} = \delta_t \qquad \forall\ t$$

The complex frequency response is

$$a(\omega)h(\omega) = 1$$

$$a(\omega) = \lim_{t\to\infty} C_t^H \underline{a}$$

$$\underline{a} = (a_0 a_1 \cdots a_{t-1})$$

Second-Order Descriptors. The correlation sequence is characterized by

$$\sum_{m=0}^{\infty} a_m \sum_{n=0}^{\infty} r_{m-n+s} a_n = \delta_s$$

from which it follows that the spectrum is

$$|a(\omega)|^2 r(\omega) = 1$$

AR(p) Case. In the AR(p) case the autoregression terminates and we may write

$$\sum_{n=0}^{p} a_n y_{t-n} = u_t$$

Obvious specializations of all first- and second-order descriptors result.

2.3 Tying Up the MA(∞) and AR(∞) Representations

Define the synthesis filter

$$H(z) = \sum_{n=0}^{\infty} h_n z^{-n}$$

and the whitening filter

$$A(z) = \sum_{n=0}^{\infty} a_n z^{-n}$$

The results of Sections 2.1 and 2.2 say that the impulse responses of H(z) and A(z) annihilate each other or, equivalently, that H(z) and A(z) whiten each other:

$$\sum_{n=0}^{\infty} a_n h_{t-n} = \delta_t$$

$$A(z)H(z) = 1$$

3.0 Linear Transformations and the Representation of Snapshots

Let Y_t denote a t-sample snapshot of (y_t) :

$$Y_t = (y_0\ y_1 \cdots y_{t-1})^H$$

The correlation matrix for Y_t is symmetric and Toeplitz:

$$R_t = \begin{bmatrix} r_0 & r_1 & & r_{t-1} \\ r_1 & & & \\ & & & r_1 \\ r_{t-1} & & & r_0 \end{bmatrix}$$

This matrix has LU Cholesky decomposition

$$R_t = H_t H_t^H$$

$$H_t = \begin{bmatrix} h_0^0 & & & & \underline{h}^0 \\ h_1^1 & h_0^1 & 0 & 0 & \\ \vdots & & & & \\ h_s^s & & h_0^s & 0 & \underline{h}^s \\ \vdots & & & & \\ h_{t-1}^{t-1} & & h_{t-1-s}^{t-1} & h_0^{t-1} & \end{bmatrix}$$

$$\underline{h}_0 \qquad \underline{h}_s$$

We call the s^{th} row vector $\underline{h}^s$ the order-s MA synthesizer, for reasons to become clear, and the s^{th} column vector $\underline{h}_s$ the impulse response of an MA linear trnasformation to an impulse applied at time s. When we tie up MA and AR theory we will see that other interpretations accrue, as well.

The matrix R_t^{-1} has UL Cholesky decomposition

$$R_t^{-1} = A_t^H A_t$$

$$A_t = \begin{bmatrix} a_0^0 & & & \\ a_1^1 & a_0^1 & & \\ \vdots & & & \\ a_s^s & a_{s-1}^s & a_0^s & \\ & & a_1^{s+1} & \\ a_{t-1}^{t-1} & & a_{t-1-s}^{t-1} & a_0^{t-1} \end{bmatrix} \begin{matrix} \\ \\ \\ \underline{a}^s \\ \\ \\ \\ \underline{a}_s \end{matrix}$$

We call the s^{th} row vector $\underline{a}^s$ the order-s MA whitener, for reasons to become clear, and the s^{th} column vector $\underline{a}_s$ the exciter of an AR linear transformation required to generate an impulsive (not impulse) response. When we tie up MA and AR theory we will find other interpretations and connections.

Whenever we speak of Y_t we have in mind the triple (Y_t, H_t, R_t) or, equivalently, the triple (Y_t, A_t, R_t^{-1}).

3.1 MA or Synthesis Transfoemation

The snapshot Y_t has MA type representation

$$Y_t = H_t U_t$$

U_t : white vector ($EU_t=0$ $EU_tU_t^H=I_t$:(txt) identity)

The output y_s may be written

$$y_s = \sum_{n=0}^{s} h_n^s u_{s-n}$$

For this reason we call $\underline{h}^s$ the order-s MA synthesizer. In the limit $s\to\infty$ we have

$$\lim_{s\to\infty} h_n^s = h_n \qquad \forall\ n\geq 0$$

where h_n is the n^{th} MA(∞) filter coefficient in the Wold decomposition.

First-Order Descriptors. Replace U_t by the identity I_t to obtain

$$Y_t = H_t$$

As I_t corresponds to a column sequence of delayed impulses, we call H_t a column sequence of impulse responses $\underline{h}_s$ to impulses applied as time s. The entry h_n^{s+n} in $\underline{h}_s$ is the response at time s+n to an impulse applied at time s. It is easy to see that

$$h_n^n = r_n/h_0^0$$

There are at least two definitions we can give for the "frequency response" of the MA transformation H_t.

LFR. The left frequency response is the object

$$H_t(\omega;L) = C_t^H(\omega)H_t$$

This is a row vector of phased (or delayed) complex frequency responses for the impulse responses $\underline{h}_s$:

$$H_t(\omega;L) = (h_0(\omega),\ldots,h_{t-1}(\omega))$$

$$h_s(\omega) = C_t^H(\omega)\underline{h}_s$$

RFR. The right frequency response is the object

$$H_t(\omega;R) = H_t\ C_t(\omega)$$

This is a column vector of phased complex frequency responses for the MA synthesizers $\underline{h}^s$:

$$H_t(\omega;R) = (h^0(\omega),\ldots,h^{t-1}(\omega))^H$$

$$h^s(\omega) = \underline{h}^s\ C_t(\omega)$$

Second-Order Descriptors. R_t is our obvious second-order descriptor. By analogy with linear filtering theory we can try to associate spectra with the norms of our complex frequency responses.

LFR. Define the left spectrum as

$$R_t(\omega;L) = |H_t(\omega;L)|^2$$

$$= C_{t(\omega)}^H R_t\ C_t(\omega)$$

$$= \sum_{s=0}^{t-1} |h_s(\omega)|^2$$

This result leads to several important observations:

(1) t^{-1} times the left spectrum is the triangular windowed or Bartlett spectrum:

$$\frac{1}{t} R_t(\omega;L) = \frac{1}{t} C_t^H(\omega)R_t\ C_t(\omega)$$

$$= \sum_{n=-(t-1)}^{t-1} \left(1-\frac{|n|}{t}\right) r_n e^{-jn\omega},$$

(2) the Bartlett spectrum is an average of magnitude-squared frequency responses for the impulse responses $\underline{h}_s$:

$$\frac{1}{t} C_t^H(\omega)R_tC_t(\omega) = \frac{1}{t}\sum_{s=0}^{t-1} |h_s(\omega)|^2,$$

(3) if the average above is (arbitrarily?) truncated at s=0, something akin to the rectangular windowed spectrum results:

$$|h_0(\omega)|^2 = \sum_{s=0}^{t-1} h_s^s e^{-js\omega}$$

$$= \frac{1}{|h_0^0|^2}\left|\sum_{n=0}^{t-1} r_n e^{-jn\omega}\right|^2$$

RFR. Define the right spectrum to be

$$R_t(\omega;R) = |H_t(\omega;R)|^2$$

$$= \sum_{s=0}^{t-1} |h^s(\omega)|^2$$

This leads to the following observations:

(1) t^{-1} times the right spectrum is an average of magnitude-squared frequency responses for the MA synthesizers $\underline{h}^s$:

$$\frac{1}{t} R_t(\omega;R) = \frac{1}{t} \sum_{s=0}^{t-1} |h^s(\omega)|^2 \quad ,$$

(2) if the average above is (arbitrarily) modified to include only the t-1 term, the following maximum order MA synthesizer spectrum results:

$$|h^{t-1}(\omega)|^2 = \left| \sum_{s=0}^{t-1} h_s^{t-1} e^{-js\omega} \right|^2$$

MA(q) Case: In the MA(q) case the MA synthesizing transformation specializes as follows:

$$H_t = \begin{bmatrix} h_0^0 & & & & \\ h_1^1 & h_0^1 & & & \\ \vdots & & & & \\ h_q^q & h_1^q & h_0^q & 0 & \underline{h}^q \\ & h_q^q & h_1^q & h_0^q & \\ & 0 & h_q^q & & h_0^q \\ & & & & h_0^q \end{bmatrix}$$

The important thing to notice here is that for $q \le s \le t-q$ $(t \ge 2q+1)$, the MA synthesizer $\underline{h}^s$ looks like the impulse response $\underline{h}_s$.

LFR. In the MA(q) case the LFR specializes as does its magnitude squared. The magnitude squared becomes

$$R_{t,q}(\omega;L) = \sum_{s=0}^{q-1} |h_s(\omega)|^2 + (t-2q)|h_q(\omega)|^2 + \sum_{s=t-1-q}^{t-1} |h_s(\omega)|^2 \quad ,$$

a linear combination that lends extra weight to the q^{th} impulse response.

RFR. In the MA(q) case the RFR specializes, as does it magnitude-squared. The magnitude-squared becomes

$$R_{t,q}(\omega;R) = \sum_{s=0}^{q-1} |h^s(\omega)|^2 + (t-q)|h^q(\omega)|^2 \quad ,$$

a linear combination that lends extra weight to the q^{th} MA synthesizer.

3.2 AR or Analysis Transformation

The snapshot Y_t has AR type representation

$$A_t Y_t = U_t$$

The outputs y_s to y_0 may be filtered to obtain

$$\sum_{n=0}^{s} a_n^s y_{s-n} = u_s$$

For this reason we call $\underline{a}^s$ the order-s MA whitener or analyzer. In the limit $s \to \infty$ we have

$$\lim_{s\to\infty} a_n^s = a_n \quad \forall \quad n \ge 0$$

where a_n is the n^{th} AR(∞) filter coefficient in the Kolmogorov representation.

First-Order Descriptors. Replace U_t by I_t to obtain

$$A_t H_t = I_t$$

As in the MA case, H_t is the column sequence of impulse responses $\underline{h}_s$. The entries in $\underline{h}_s$ have the same interpretation as before.

If U_t is replaced by an exciter matrix that makes $Y_t = I_t$, an impulsive response, we obtain

$$A_t = E_t$$

For this reason we call A_t also a column sequence of impulsive exciters.

LFR. The left frequency response is the object

$$A_t(\omega;L) = C_t^H(\omega) A_t$$

This is a row vector of phased complex frequency responses for the impulsive exciters $\underline{a}_s$:

$$A_t(\omega;L) = (a_0(\omega), \ldots, a_{t-1}(\omega))$$

$$a_s(\omega) = C_t^H(\omega) \underline{a}_s$$

RFR. The right frequency response is

$$A_t(\omega;R) = A_t C_t(\omega)$$

This is a column vector of phased complex frequency responses for the MA whiteners $\underline{a}^s$:

$$A_t(\omega;R) = (a^0(\omega) \ldots a^{t-1}(\omega))^*$$

$$a^s(\omega) = \underline{a}^s C_t(\omega)$$

Second-Order Descriptors. R_t (or R_t^{-1}) is our obvious second-order descriptor. We can try to associate spectra with the norms of the complex frequency responses. But, as A_t is a whitening transformation, we ought to use the inverse of these spectra as our spectral definitions for Y_t.

LFR. Define the left whitening or analysis spectrum as

$$R_t^-(\omega;L) = \frac{1}{|A_t(\omega;L)|^2} = \frac{1}{C_t^H(\omega) A_t A_t^H C_t(\omega)} = \frac{1}{\sum_{s=0}^{t-1} |a_s(\omega)|^2}$$

Several observations can be made:

(1) The inverse of t times the left whitening spectrum is an average of order-increasing

exciter spectra:

$$\frac{1}{tR_t^-(\omega;L)} = \frac{1}{t}\sum_{s=0}^{t-1}|a_s(\omega)|^2$$

(2) If the sum is truncated at s=0, the resulting spectrum is the inverse of the exciter spectrum $|a_0(\omega)|^2$.

RFR. Define the right whitening spectrum as

$$R_t^-(\omega;L) = \frac{1}{|A_t(\omega;R)|^2}$$

$$= \frac{1}{C_t^H(\omega)R_t^{-1}C_t(\omega)}$$

$$= \frac{1}{\sum_{s=0}^{t-1}|a^s(\omega)|^2}$$

Several interesting observations may be made:

(1) The right whitening spectrum is the maximum likelihood method (MLM) spectrum,

(2) The inverse of t times the right whitening spectrum is an average of order-increasing whitening spectra:

$$\frac{1}{tR_t^-(\omega;R)} = \frac{1}{t}\sum_{s=0}^{t-1}|a^s(\omega)|^2$$

(3) if the sum is arbitrarily modified to include only the t-1 term, the maximum entropy method (MEM) spectrum results:

$$\frac{1}{|a^{t-1}(\omega)|^2}$$

AR(p) Case. In the AR(p) case the AR synthesizing transformation specializes as follows:

$$A_{t,p} = \begin{bmatrix} a_0^0 & & & & \\ a_1^1 & a_1^0 & & 0 & \\ \vdots & & \ddots & & \\ a_p^p & \cdots & a_0^p & & \\ & a_p^p & & a_0^p & \\ & & \ddots & & \\ & & a_p^p & & a_0^p \end{bmatrix}$$

For $p \le s \le t-p$ $(t \ge 2p+1)$ the MA whitener $\underline{a}^s$ looks like the transpose of the exciter $\underline{a}_s$.

LFR. In the AR(p) case the LFR specializes as does its magnitude-squared. The left whitening spectrum becomes

$$R_{t,p}^-(\omega;L) = \frac{1}{\sum_{s=0}^{p-1}|a_s(\omega)|^2+(t-2q)|a_p(\omega)|^2+\sum_{s=t-1-q}^{t-1}|a_s(\omega)|^2}$$

a linear combination that lends extra weight to the p^{th} exciter spectrum.

RFR. In the AR(p) case the right whitening spectrum specializes to the following linear combination:

$$R_{t,p}^-(\omega;R) = \frac{1}{\sum_{s=0}^{p-1}|a^s(\omega)|^2+(t-p)|a^p(\omega)|^2}$$

This lends extra wight to the p^{th} MA whitener.

3.3 Tying up the MA and AR Theory

We have the relations

$$A_t H_t = I_t$$

$$H_t A_t = I_t$$

From these the most succinct interpretations of H_t and A_t are the following:

$$H_t = \begin{bmatrix} \underline{h}^0 & & & \\ \underline{h}^1 & & & \\ \underline{h}^2 & & 0 & \\ & \ddots & & \\ & & & \underline{h}^{t-1} \end{bmatrix}$$; $\underline{h}^s$:order-s MA synthesizer

$$= \begin{bmatrix} \underline{h}_0 & & & \\ \underline{h}_1 & & & \\ \underline{h}_2 & & 0 & \\ & \ddots & & \\ & & & \underline{h}_{t-1} \end{bmatrix}$$; $\underline{h}_s$:impulse response for AR(s) filter corresponding to order-s MA whitener $\underline{a}^s$

$$A_t = \begin{bmatrix} \underline{a}^0 & & & \\ \underline{a}^1 & & & \\ \underline{a}^2 & & 0 & \\ & \ddots & & \\ & & & \underline{a}^{t-1} \end{bmatrix}$$; $\underline{a}^s$:order-s MA whitener or analyzer

$$= \begin{bmatrix} \underline{a}_0 & & & \\ \underline{a}_1 & & & \\ \underline{a}_2 & & & \\ & & & \underline{a}_{t-1} \end{bmatrix}$$; $\underline{a}_s$:impulse response for AR(s) filter corresponding to order-s MA synthesizer $\underline{h}^s$

4.0 Parametrized Spectra

The results for MA(q) and AR(p) linear transformations suggest that the MA synthesis and the AR analysis transformations H_t and A_t may be approximated by the q^{th} and p^{th} order approximants $H_{t,q}$ and $A_{t,p}$. This kind of thinking in the linear transformation world is directly analogous to the kind of thinking that goes on in the linear filtering world when we identify order-q MA and order-p AR filters to model data that is surely infinite-dimensional. So think of $H_{t,q}$ and $A_{t,p}$ as parametrized synthesis and analysis linear transformations that can be used in place of H_t and A_t in the various spectra defined in Sections 3.1 and 3.2 to obtain parametrized analysis and synthesis spectra.

In the sections to follow we explore the so-called parametrized maximum likelihood method (pMLM) spectrum that results from replacing A_t by $A_{t,p}$ in the right whitening spectrum:

$$R^{-}_{t,p}(\omega;R) = \frac{1}{C_t^H R_{t,p}^{-1} C_t}$$

$$= \frac{1}{\sum_{s=0}^{p-1} |a^s(\omega)|^2 + (t-p)|a^p(\omega)|^2}$$

$$R^{-1}_{t,p} = A^H_{t,p} A_{t,p}$$

Let's agree to call $R^{-}_{t,p}(\omega R)$, the pMLM spectrum $MLM_{t,p}(\omega)$.

4.1 Limiting Results

The pMLM spectrum may be written

$$\frac{1}{MLM_{t,p}(\omega)} = \frac{1}{MLM_{p-1}(\omega)} + \frac{t-p}{MEM_p(\omega)}$$

where $MLM_{p-1}(\omega)$ is the maximum likelihood method spectrum of order p-1 and $MEM_p(\omega)$ is the maximum entropy method spectrum of order p.

p=t-1. In this case the pMLM spectrum is the MLM_{t-1} spectrum.

p fixed; t↑∞. In this case the pMLM spectrum goes to the MEM_p spectrum:

$$\lim_{t\uparrow\infty} MLM_{t,p}(\omega) = MEM_p(\omega)$$

From these results we see that the pMLM spectrum defines a two-parameter class whose properties lie, in some sense, between those of the MLM and the MEM spectra.

4.2 Numerical Results

We examine the performance of the pMLM spectrum for three different sets of correlation data. For purposes of comparison with published results, we use the example of two close sine waves in noise studied by Lacoss :

$$r_t = \delta_t + 5.33\cos(.3\pi t) + 10.66\cos(.4\pi t) \quad t\in[0,21].$$

In Figure 1 it is observed that by extending the correlation matrix to obtain the $MLM_{21,11}(\omega)$ spectrum we resolve quite successfully the two close peaks while the $MLM_{11}(\omega)$ cannot. Another interesting observation concerns the power estimation. The parametrized $MLM_{21,11}(\omega)$ estimates accurately the 3dB difference in the peaks, while the $MEM_{11}(\omega)$ spectrum gives on the order of 5dB.

The second example shown in Figure 2 is based on synthetic AR data. We generate 4th order AR data corresponding to the transfer function

$$H(z) = \frac{1}{1-.8z^{-4}}$$

The same comments can be made - the parametrized $MLM_{21,5}(\omega)$ yields a much better resolution than the $MLM_5(\omega)$.

Finally in Figure 3 we have applied the method to a finite set of recorded speech data. The data used is a set of 80 data points sampled from the vowel "I" in the sentence, "I hope it's April". This example illustrates very well the compromise between resolution and smoothness. The three spectra illustrated require exactly the same amount of computation. However, the parametrized $MLM_{21,11}(\omega)$ exhibits a dramatic improvement in resolution over the $MLM_{11}(\omega)$ and a much better power estimation of the second peak over the $MEM_{11}(\omega)$.

The results are summarized in the following way. The pMLM spectrum can be used in two conceptually different ways. First if one is interested in saving parameters, then a resolution comparable to the $MLM_t(\omega)$ can be achieved with the $pMLM_{t,p}$ with p<t.

On the other hand if better resolution is sought $\{r_t\}_0^{t_0}$ is available, then by extending the correlation matrix to $t>t_0$, a resolution close to the $MEM(t_0)$ can be obtained while preserving the power estimation of the MLM. As a result we see that the parametrized MLM realizes a compromise between the smoothness of the MLM spectrum and the resolution of the MEM.

4.3 Order Fitting

The essential problem in all of this is order-fitting: deciding when $R_{t,p}$, for example, is close enough to R_t to decide that $H_{t,p}$ or $A_{t,p}$ is a good enough approximation to H_t or A_t. In the context of the parametrized maximum likelihood method (pMLM) two order fitting rules arise naturally.

J-Divergence: Suppose we are given a finite data set Y_t with correlation matrix R_t. The parametric model is determined by the correlation matrix $R_{t,p}$ as defined earlier. The problem of order fitting is one of discriminating between the two hypotheses H_t and $H_{t,p}$ from the data Y_t:

$$H_t : Y_t : \quad N(0,R_t)$$

$$H_{t,p} : Y_t : \quad N(0,R_{t,p}) \quad .$$

The divergence between the two hypotheses is defined as:

$$J_t(p,t) = E_{H_{t,p}} L_t(Y_t) - E_{H_t} L_t(Y_t)$$

where $L_t(x)$ is the likelihood function $L_t(x) =$

$P(x/H_{t,p})/P(x/H_t)$. The set of data is supposed normally distributed. Using the identity $E(x^TQx) = Tr(QE(xx^T))$, the J-divergence is

$$J_t(p,t) = \frac{1}{2t} Tr[R_t R_{t,p}^{-1} + R_{t,p} R_t^{-1} - 2I].$$

The J-divergence, while not a true metric, provides a measure of the distance between the two hypotheses. It can also be viewed as a measure of how far $R_t R_{t,p}^{-1}$ or $R_{t,p} R_t^{-1}$ are from the identity matrix. If λ_i, i=0,1,..., t-1, are the eigenvalues of the matrix $R_t R_{t,p}^{-1}$, then we can also write

$$J_t(p,t) = \frac{1}{2t} \sum_{i=0}^{t-1} (\sqrt{\lambda_i} - \frac{1}{\sqrt{\lambda_i}})^2 \quad .$$

This suggests that the J-divergence is more simply interpreted as a measure of the dispersion of the eigenvalues around unit.

J-divergence is twice the Kullback-Leiblerinformation used by Akaike in his discussion of information criteria and order-fitting. Thus Akaike's derivation of the AIC can be translated in its integrity in terms of the J-divergence, and the following order fitting rule is proposed:

$$J_t(p) = J_t(t,p) + 2\frac{p}{t}$$

A Simpler Distance Measure: The calculation involved in the J-divergence is quite large. We are after a simple order fitting rule easy to compute at each order p. An intuitive measure of the distance between R_t^{-1} and $R_{t,p}^{-1}$ consists in getting a measure of how far $\frac{1}{a_i}$, p<i<t-1, are from the constant value $\frac{1}{a_p}$. Consider the following quadratic mean:

$$DME(p) = \frac{1}{t-p-1} \sum_{i=p-1}^{t-1} (\frac{1}{a_i} - \frac{1}{a_p})^2$$

Note that if the given correlation sequence corresponds to an exact autoregressive process of order p_0, then DME(p)=0 for $p \geq p_0$.

Numerical Results: In Figure 4 we apply the criteria to the set of correlation data estimated from synthetic AR(4) data generated in Example n°2 of Section 4.2. The simple measure and the J-divergence we compare to the minimum prediction error. Figure 5 illustrates the similarity of behavior between the J-divergence corrected for asymptotic bias and the AIC criterion. We have applied the criteria to the speech data in Figure 6. Again the J-divergence corrected behaves like the AIC but has a sharper and clearer minimum. Here the simple measure has a slow convergence rate that would be misleading in the order fitting.

5.0 Conclusions

Many of the ideas in this paper are speculative. The spectra defined seem to be plausible definitions for the frequency responses of nonstationary synthesizers and whiteners (analyzers). The fact that the argumentation leads to Bartlett and MLM spectra lends some credence to the line of reasoning. The numerical results indicate that pMLM - and hopefully its other parametrized counterparts as well - can become useful adjuncts to MLM and MEM processing. The pMLM seems to offer a tradeoff between the resolution of MEM and the power fidelity of MLM.

All discussions of MA(q) and AR(p) linear transformation generalizes to ARMA(p,q) linear transformation. Such transformations underly lattice and innovations representations of stationary time series and the corresponding lattice and Kalman predictor structures that play such an important role in exact likelihood for ARMA time series. The nonstationarity (or initial conditions) of the linear transformations are captured in the order increase for the lattice and the time variation for the Kalman gain. By combining fast Levinson/Durbin and impulse response algorithms, the lattice and Kalman predictors lead to fast and exact computation of exact likelihood.

Koopmans [1] contains a more complete discussion of the MA(∞) and AR(∞) theory of stationary time series. The interpretations of R_t and R_t^{-1} are evident in the influential work of Friedlander, Kailath, and Morf. The connectionsbetween MLM and MEM are due to Burg [2]. The interpretation we give for the Bartlett spectrum may be close to that given by Oppenheim at the 1980 l'Aquila conference. The Lacoss example is reported in [3]. The appropriate references to Akaike are [4] and [5]. Other topical work on order-fitting is reported by Parzen [6] and Rissanen [7]. Our motivation to use J-divergence arose from [8] and [9]. The simplified DME rule is discussed more fully in [10]. Exact likelihood is reported in [10]-[12]. Variations on and fast versions of the Levinson-Durbin algorithm [13]-[14] are reported in [15], [16], and [10], where respectively the MSK algorithm, the impulse response algorithm, and a fast impulse response algorithm are derived.

References

[1] L. J. Koopmans, "The Spectral Analysis of Time Series," Academic Press, New York, 1974.

[2] J. P. Burg, "The Relationship between Maximum Entropy Spectra and Maximum Likelihood Spectra," Geophysics, Vol. 37, April 1972.

[3] R. T. Lacoss, "Data Adaptive Spectral Analysis Methods," Geophysics, Vol. 26, 4, Aug. 1971.

[4] H. Akaike, "A New Look at the Statistical Model Identification," IEEE Trans. on Aut. Cont., AC-19, Dec. 1974.

[5] H. Akaike, "Statistical Predictor Identification," Ann. Inst. Stat. Math., Vol. 22, 1970.

[6] E. Parzen, "Some Recent Advances in Time Series Modelling," IEEE Trans. on Aut. Cont., AC 19, 1974.

[7] J. Rissanen, "Modelling by the Shortest Data Description," Automatica, Vol. 14, 1978.

[8] S. Kullback, "Information Theory and Statistics," Dover Pub., New York, 1968.

[9] L. L. Scharf and P. H. Moose, "Information Measures and Performance Bounds for Array Processors," IEEE on Inf. Theory, Vol. IT-22, Jan. 1976.

[10] J. P. Dugre, "Parametric Spectrum Analysis for Stationary Random Sequences," Ph.D. dissertation, Colorado State University, 1981.

[11] R. H. Jones, "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations," Technometrics, Aug. 1980.

[12] C. Gueguen and L. L. Scharf, "Exact Likelihood Identification of ARMA Models, A Signal Processing Perspective," Proc. EUSIPCO, Lausanne, Switzerland, Sept. 1980.

[13] N. Levinson, "The Wiener RMS (root-mean-square) Error Criterion in Filter Design and Prediction," J. Math. Phys., Vol. 25, Jan. 1947.

[14] J. Durbin, "The Fitting of Time Series Models," Rev. Int. Stat. Inst., Vol. 28, 1960.

[15] M. Morf, G. S. Sidhu and T. Kailath, "Some new algorithms for Recursive Estimation on Constant Linear Discrete Time Systems," IEEE Trans. on Aut. Cont., AC-19, August 1974.

[16] J. Leroux and C. Gueguen, "A Fixed Point Computation of the Partial Correlation Coefficients," Trans. on ASSP, Vol-25, n°2, June 1977.

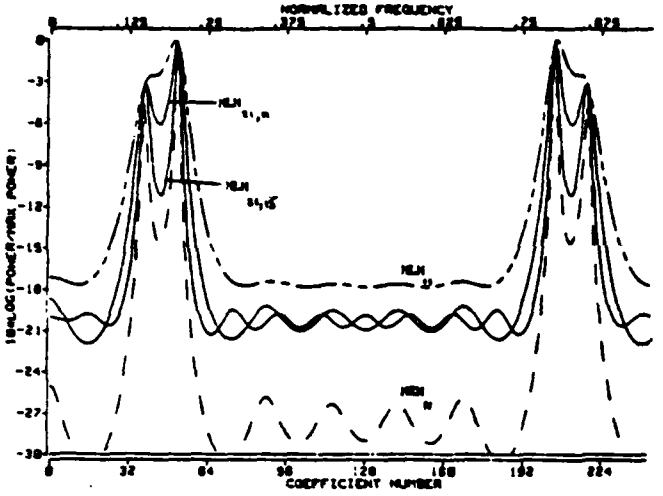


Figure 1 Close Sine Waves in Noise
$r_t = \delta_t + 5.33 \cos(.3\,\pi t) + 10.66 \cos(.4\,\pi t)$

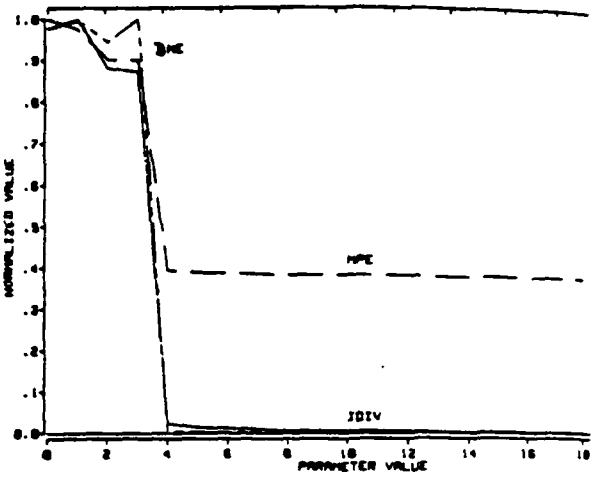


Figure 4 Synthetic AR Data
$H(z) = 1/(1 - .8\,z^{-4})$

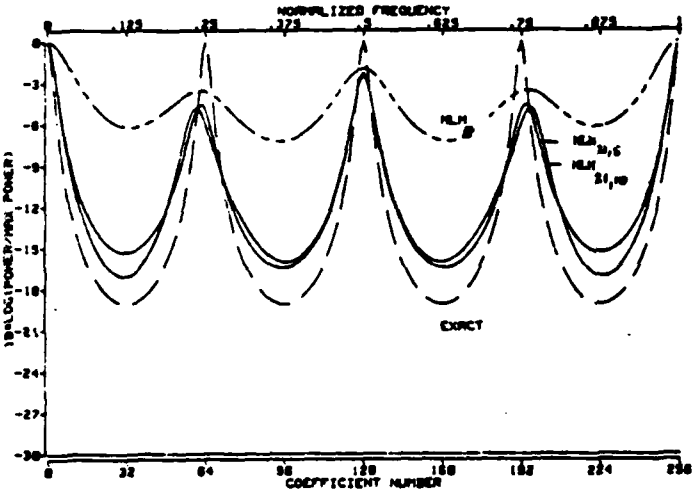


Figure 2 Synthetic AR Data
$H(z) = 1/(1 - .8\,z^{-4})$

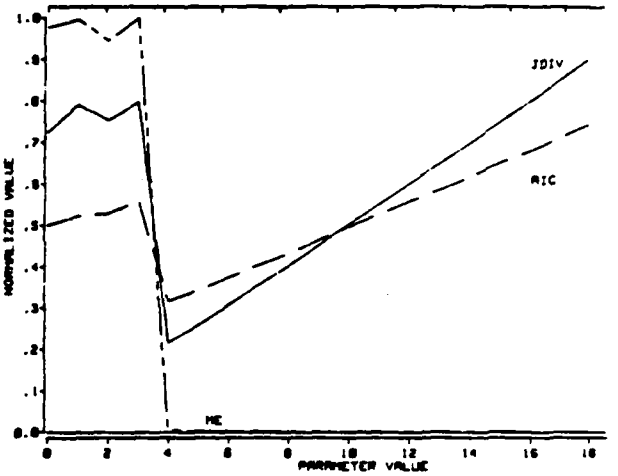


Figure 5 Synthetic AR Data

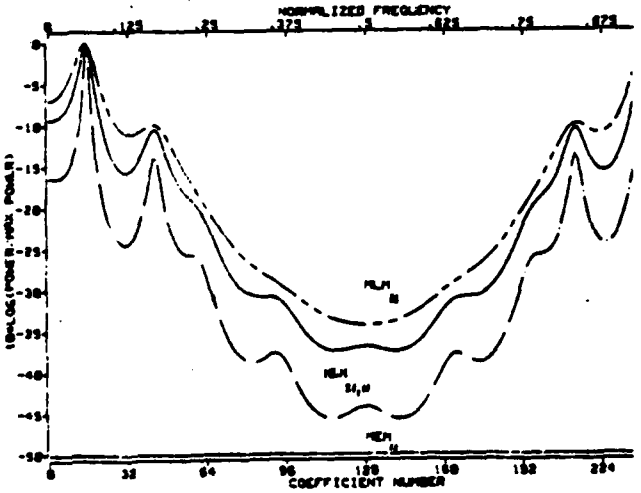


Figure 3 Speech Data
Number of data points: 88
Sample interval: .1 ms

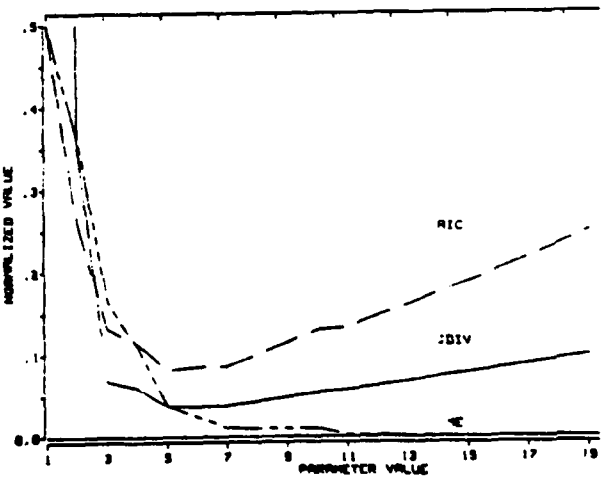


Figure 6 Speech Data

EXACT LIKELIHOOD FOR STATIONARY VECTOR AUTOREGRESSIVE MOVING AVERAGE PROCESSES

Jean-Pierre Dugre
Louis L. Scharf
Claude J. Gueguen

Presented at:

Workshop on Fast Algorithms in Linear Systems
Aussois, France
September 1981

EXACT LIKELIHOOD FOR STATIONARY VECTOR AUTOREGRESSIVE MOVING AVERAGE PROCESSES[1]

Jean-Pierre Dugré
Electrical Engineering Department, Colorado State University
Fort Collins, CO 80523

Louis L. Scharf
Electrical Engineering Department, Colorado State University
Fort Collins, CO 80523

Claude Gueguen
Ecole Nationale Superieure de Telecommunications
46 rue Barrault 75634 Paris

Abstract

Les modèles autoregressifs à moyenne mobile (ARMA) sont des approximations utiles des processus aléatoires communement rencontrés dans le traitement des signaux à temps discrets. De tels modèles sont utilisés à la compression des données en transmission d'informations à bas débit, à l'amélioration de la résolution en analyse spectrale, à la prévision en économie, météorologie et autres series numériques.

Dans cet article nous discutons differents aspects de l'identification des modèles ARMA par le maximum de vraissemblance. Nous soulignons le rôle de la représentation de "l'innovation" dans le calcul de la fonction de vraissemblance exacte. Finalement nous montrons comment la structure interne du modèle peut être mise à profit pour accélérer le calcul de la fonction de vraissemblance soit par un prédicteur rapide de Kalman soit par l'implémentation d'une structure en échelle (Lattice) rapide.

Autoregressive-moving average (ARMA) models, are useful approximants to the kinds of random processes commonly encountered in discrete-time signal processing applications. Such models may be used to compress data in low bit-rate information transmission, improve frequency resolution in spectrum analysis, and to forecast in economic, meteorological, and other time series.

In this paper we discuss several aspects of the maximum likelihood theory of parameter identification in ARMA models. We highlight the role of innovations representations in exact likelihood theory and show how internal model structure may be used to speed up calculation of likelihood in either fast Kalman predictor or fast lattice implementations.

[1]This work supported by the Office of Naval Research, Statistics and Probability Branch, Arlington, VA 27740, under contract N00014-75-C-0518, and by the Army Research Office, Research Triangle Park, NC 27709, under contract DAAG29-79-C-0176.

1. INTRODUCTION

The random processes encountered in signal processing applications are typically lowpass or bandpass processes in which redundancy is high. This means finite-dimensional models may often be used to approximate the second-order properties of the processes. The dominant motivations for using finite-dimensional models are (1) they provide a systematic framework for deriving data compression and frequency resolution improving algorithms, and (2) they become predictor formulae for event forecasting.

The problems of data compression, resolution improvement, and forecasting are "solved", so to speak, by identifying a parametric model that either infinitely extends a data correlation sequence or matches the data, itself, in a least squares or maximum likelihood sense.

AR models suffer the defect that spectral zeros are not easily modeled with low-order schemes. Couple to this defect the fact that sample-data versions of rational continuous-time processes are autoregressive moving average (ARMA), and we have strong motivation for identifying the more general ARMA models.

Traditionally the emphasis in identification of ARMA models has been on approximate representations (such as "long ARs") that lead to linear identification procedures. However, more recently there has been a flurry of activity in exact maximum likelihood formulations and nonlinear optimization procedures.

[Box and Jenkins, 1976] developed the familiar conditional sum of squares for the identification of univariate MA processes and treated the ARMA case as a special MA of infinite dimension. This method was later generalized by [Newbold, 1974] and [Ali, 1977] to mixed processes. They obtained expressions for the inverse and determinant of the sample correlation matrix from which the exact likelihood could be computed. [Osborn, 1976] applied the same approach to the case of multivariate moving average processes.

Using a different type of linear transformation of the input white noise sequence, [Phadke and Kedem, 1978] showed how to obtain exact likelihood

for a pure moving average process by Cholesky decomposition of the correlation matrix. This approach was generalized to the mixed models case by [Ansley, 1979], who presented the first really efficient algorithm.

[Akaike, 1973] and [Anderson, 1978] have tried to obtain an analytical solution to the problem of exact maximum likelihood for vector ARMA processes. [Akaike, 1973] formulates the problem directly as the identification of a Gaussian model by numerical maximization of the Gaussian likelihood function. Following [Kashyap, 1970], he concentrates on obtaining expressions for the gradient and Hessian of the log likelihood function to be used in a Newton-Raphson type non-linear procedure. All these methods are in fact approximate in the sense that they consider the conditional likelihood with conditioning on some fixed initial conditions. [Anderson, 1978] and later [Arhabi, 1978, 1979] further developed the method and expressed the likelihood function in terms of the ratio of the periodogram to the spectral density function of the model. Along these lines, the following work is also relevant: [Tretter and Steiglitz, 1967], [Tunnicliffe Wilson, 1973], [Dunsmuir and Hannan, 1976], [Shaman, 1973], [Galbraith and Galbraith, 1974].

All the preceeding methods can be regarded as computationally impractical. The success of the maximum likelihood theory as an identification procedure for ARMA processes is directly tied up with the ability to efficiently compute the likelihood function.

An alternative representation of an ARMA process is the Markovian representation, introduced by [Akaike, 1974]. As early as 1965, [Schweppe, 1965] had indicated how Kalman filtering theory could be used to get the exact likelihood function in the scalar case. Later [Harvey and Phillips, 1979] further developed the theme. The method has been adapted to processes with missing data in a very useful paper by [Jones, 1980]. Recently [Gueguen and Scharf, 1980], using a somewhat different approach, based on the innovations representation of an ARMA time series, have given new and interesting filtering interpretations. These interpretations show the connection between Markovian and innovations representations of a time series and show how the Kalman gain vector is related to the impulse responses of increasing order autoregressive models fitted on the data.

In this paper we generalize these concepts to vector autoregressive moving average models and obtain efficient recursive procedures to compute the exact likelihood function in fast Kalman predictor and fast lattice forms.

2. EXACT LIKELIHOOD FOR VECTOR ARMA PROCESSES

We assume that a finite set of observations $\{y_0^T,\ldots,y_N^T\}$ on a d dimensional, zero mean, wide sense stationary random process $\{y_t\}$ is given. The problem to solve is one of fitting a vector ARMA (p,q) model, $q \leq p$,

$$\sum_{m=0}^{p} a_m\, y_{t-m} = \sum_{n=0}^{q} b_n\, u_{t-n} \qquad a_0 = 1, \text{ for all } t \tag{1}$$

to the data. $\{u_t\}$ is an input d dimensional, zero mean Gaussian white noide process whose (dxd) correlation matrix is $E(u_t u_t^T) = W$.

The likelihood function is defined as the joint probability density of the set of vector data $(y_0^T,\ldots,y_N^T)$, evaluated at the observations and parameterized by the ARMA parameters. Define the following ((N+1)dx1) observation vector

$$\underline{y}^T = [y_0^T,\ldots,y_N^T] \quad .$$

The vector $\underline{y}$ is distributed as a multivariate N(0,R) where R is an (N+1)dx(N+1)d block Toeplitz matrix:

$$R = \begin{bmatrix} r_0 & r_1 & \cdots & r_N \\ r_{-1} & r_0 & & \\ \vdots & & \ddots & \\ r_{-N} & & & r_0 \end{bmatrix} \qquad r_t = E[y_t y_0^T] = r_{-t}^T$$

We shall derive several expressions for the likelihood function: conventional, Markovian, and innovations.

2.1. Conventional representation of the likelihood. The (N+1)dx(N+1)d matrix R is block Toeplitz symmetric but is not symmetric Toeplitz. The likelihood function takes the form of the well known multivariate normal density

$$L(\underline{y}) = \frac{1}{(2\pi)^{(N+1)/2}} |R|^{-1/2} \exp\{-\frac{1}{2} \underline{y}^T R^{-1} \underline{y}\} \quad (2)$$

and the log-likelihood reduces to

$$\log(L(y)) = -\frac{(N+1)}{2} \log 2\pi - \frac{1}{2} \log |R| - \frac{1}{2} \underline{u}^T \underline{u} \quad (3)$$
$$\underline{u} = R^{-1/2} \underline{y}$$

This is the vector closed-form often encountered in the literature [Ansley, 1979], [Anderson, 1978]. It is highly non-linear in terms of the ARMA parameters $(a_1,\ldots,a_p,b_0,\ldots,b_q,W)$. Several approaches can be considered to maximize this expression. The conventional direct method consists in deriving an approximate version of the likelihood by conditioning on some initial conditions.

2.2. Markovian representation of the likelihood. An alternative approach is to use the structure introduced in the Markovian representation of the process $\{y_t\}$. A stochastic process $\{y_t\}$ is said to exhibit a Markov property if the future behavior of the process can completely be described by some present state and the future input. The state condenses all the information of the present and past of the process $\{y_t\}$. Assuming the process is stable and minimum phase, [Akaike, 1974] has established the equivalence of the Markovian and ARMA representations. The finiteness of the dimension of the predictor space of an ARMA process is the fundamental characteristic of a process with a Markovian representation.

Internal structure. The finite basis of the predictor space is chosen as the state of the process. The process $\{y_t\}$ has the infinite MA representation

$$y_t = \sum_{k=0}^{\infty} h_k u_{t-k} \quad (4)$$

where $u_t = y_t - y_t\big|_{t-1}$ is the innovation of $\{y_t\}$ at time t. The analysis of the different predictors $y_t\big|_{t-1}$, $y_{t+1}\big|_{t-1}$, ..., reveals that they are given by

$$y_{t+s}\big|_{t-1} = L(y_t\big|_{t-1},\ldots,y_{t+p-1}\big|_{t-1},u_t) \qquad s = 1,2,\ldots$$

where L represents a linear transformation. The predictor space is then of finite dimension p and the vector of one-step ahead to p-step ahead predictions forms a basis for the predictor space. The state space model we obtain in the following paragraphs is based on the concept of a finite-dimensional predictor space. It is slightly different from the one given by [Akaike, 1974] since our definition of the state is a shifted version of his.

Notice from the infinite MA representation (4) that the predictors are written,

$$y_{t+i+1}\Big|_t = y_{t+i}\Big|_{t-1} + h_i\, u_{t-1} \qquad \text{for all } i \geq 0. \tag{5}$$

From (1), the subsequent relation also holds:

$$y_{t+p}\Big|_t = -\sum_{i=1}^{p} a_i\, y_{t+p-i}\Big|_{t-1} + h_p\, u_{t-1} \quad . \tag{6}$$

These recursions impose an internal structure on the process. That is, if we define the (dpx1) state vector $X_t^T = [y_t^T\Big|_{t-1}, \ldots, y_{t+p-1}^T\Big|_{t-1}]$, we obtain

$$X_t = A\, X_{t-1} + B\, u_{t-1}$$

The vector $u_t = y_t - y_t\Big|_{t-1}$ is the innovation of the process as defined earlier and has a multivariate normal density with zero mean and covariance matrix W. A and B are respectively a (dpxdp) block state matrix and a (dpxd) input matrix, defined as follows:

$$A = \begin{bmatrix} 0 & I & & \\ & & \ddots & 0 \\ & & & \ddots \\ 0 & & & I \\ -a_p & \cdots & & -a_1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} h_1 \\ \vdots \\ h_p \end{bmatrix}$$

Hence the stationary stochastic process $\{y_t\}$ is represented as a linear map from the input innovation process $\{u_t\}$ to $\{y_t\}$ via a Markovian state space structure

$$X_t = A\,X_{t-1} + B\,u_{t-1}$$

$$y_t\Big|_{t-1} = C^T X_t \qquad (7)$$

$$y_t = C^T\,X_t + u_t$$

with C^T defined as the output (dxdp) matrix [I 0 . . . 0].

Initial conditions. The wide-sense stationarity of the process $\{y_t\}$ imposes the condition that at time t=0 steady state has been achieved. This means the time origin is rejected back to t=-∞ and the initial state X_0 is a zero mean (dpx1) random vector with probability density $N(0,Q_0)$. The (dpxdp) covariance matrix Q_0 satisfies the Lyapunov equation

$$Q_0 = A\,Q_0\,A^T + B\,W\,B^T \quad .$$

First order descriptor. The first order descriptor of the vector process is the response to a unit pulse δ_t applied as the i^{th} element of the vector input u_t. The resulting output corresponds to the i^{th} column of the (dxd) impulse response matrix h_t. The impulse response matrix is directly obtained by replacing the input vector u_t by a (dxd) input diagonal matrix Δ_t, [Wolovich, 1974]

$$\Delta_t = \mathrm{diag}\,(\delta_t,\ldots,\delta_t)$$

$$\delta_t = \begin{cases} 1 & t = 0 \\ 0 & t \neq 0 \end{cases} \quad .$$

The first order descriptor is then

$$\begin{aligned} h_t &= 0 && \text{for all } t < 0 \\ h_0 &= I && t = 0 \\ h_t &= C^T A^{t-1} B && \text{for all } t > 0 \end{aligned} \qquad (8)$$

Second order descriptor. The lagged correlation matrices for the random process $\{y_t\}$ are defined by $r_t = E(y_t y_0^T)$. Using the above Markovian representation we obtain

$$r_t = E[(C^T X_t + u_t)(C^T X_0 + u_0)^T]$$
$$= C^T A^t Q_0 C + C^T A^{t-1} B W \quad . \tag{9}$$

That is, the correlation matrix for lag t is

$$r_t = C^T A^t Q_0 C + h_t W \quad . \quad t > 0$$

These equations are straightforward generalizations of the scalar version [Gueguen/Scharf, 1980].

Likelihood. In this representation, the stationary sequence of prediction residuals $\{u_t\}$ may be written

$$u_t = y_t - y_t\Big|_{t-1} \quad ,$$

This sequence is a sequence of i.i.d. random vectors with normal distribution, N(0,W). The joint distribution of the input vector $\underline{u}^T = (u_0^T, \ldots, u_N^T)$ is then readily written as the product of normal distributions,

$$f(u_0,\ldots,u_N) = \prod_{t=0}^{N} N_{u_t}(0,W) \quad .$$

These residuals may be computed causally from the time series values $\underline{y}^T = (y_0^T,\ldots,y_N^T)$ provided the initial state is given:

$$u_t = y_t - y_t\Big|_{t-1}$$
$$y_t\Big|_{t-1} = C^T X_t$$
$$X_t = A X_{t-1} + B u_{t-1}$$
$$y_0\Big|_{-1} = C^T X_0 \quad .$$

Thus we may write the conditional likelihood function

$$L(\underline{y}/X_0) = \prod_{t=0}^{N} f(u_t + y_t\Big|_{t-1}/X_0) = \prod_{t=0}^{N} N_{y_t}(y_t\Big|_{t-1}, W) \quad . \tag{10}$$

The exact likelihood function of the process is then obtained by integrating over all the realizations of the random initial state X_0 which has a

normal density $N(0,Q_0)$. Then

$$L(\underline{y}) = \int_{\text{all } X_0} \prod_{t=0}^{N} N_{y_t}\left(y_t\Big|_{t-1}, W\right) N_{X_0}(0,Q_0)\, d X_0 \quad . \tag{11}$$

This result generalizes the scalar result of [Gueguen/Scharf, 1980]. It is of theoretical interest as we shall see later, but of no computational value. A maximum likelihood identification procedure based on the Markovian representation could be derived using the conditional likelihood (10). An interesting problem then is one of choosing the appropriate initial state X_0 [Gueguen/Scharf, 1980].

2.3. The Innovations Representation of Likelihood. [Gueguen and Scharf, 1980], drawing on ideas from [Anderson and Moore, 1978], showed how a linear time varying predictor, or innovations, representation could be derived from the Markovian representation. The essential advantage of such a representation is that the initial predictor state is no longer a random vector, but is rather an identically zero vector. These ideas generalize to the multivariate case. This feature is of inestimable value in likelihood theory, as we shall see.

Internal Structure. We are looking for a model that will produce an output y_t that has the same statistical properties as the original ARMA process. The innovations model maps, some time-varying zero-mean, Gaussian white-noise $\{\hat{u}_t\}$ into $\{y_t\}$ through a time varying structure. The states are still defined as a (dpx1) vector of predictors, $\hat{X}_t^T = [\hat{y}_t^T\Big|_{t-1} \ \ldots \ \hat{y}_{t+p-1}^T\Big|_{t-1}]$, but now they evolve according to the time varying state equation,

$$\hat{X}_t = A\,\hat{X}_{t-1} + k_{t-1}\,\hat{u}_{t-1} \quad . \tag{12}$$

Here k_t is a (dpxd) time varying Kalman block vector that replaces the time invariance B of the Markovian representation:

$$k_t^T = [k_{1,t}^T \ \ldots \ k_{p,t}^T] \quad ,$$

and $\{\hat{u}_t\}$ is a multivariate nonstationary $N(0,w_t)$ innovation or prediction residual that replaces the stationary residual sequence $\{u_t\}$ in the

Markovian representation. $w_t = E[\hat{u}_t \hat{u}_t^T]$ is a (dxd) zero-lag correlation matrix. The output is then given by the Markovian measurement equation

$$y_t = \hat{y}_t\Big|_{t-1} + \hat{u}_t = C^T \hat{X}_t + \hat{u}_t \quad .$$

Initial conditions. The origin of time may be brought back to t=0 and the initial conditions may now be given deterministically as

$$\begin{aligned} X_0 &\equiv 0 \\ Q_0 &= E[X_0 X_0^T] \equiv 0 \quad . \end{aligned} \tag{13}$$

First order descriptor. Substituting the diagonal matrix Δ_t in place of the vector $\hat{u}_t$ leads to the following expression for the impulse response:

$$\begin{aligned} h_i^t &= 0 && \text{for all } i < 0 \\ h_i^t &= I && i = 0 \\ h_i^t &= C^T A^{i-1} k_t && \text{for all } i > 0 \quad . \end{aligned} \tag{14}$$

The term h_i^t is the response of the representation (12) at time t+i to an input Δ_t that applies a diagonal input at time t.

Second order descriptor. The (dxd) matrix correlation matrix sequence $\{r_i^t\}$ of the output process is readily obtained as follows:

$$\begin{aligned} r_i^t &= E[y_{t+i} y_t^T] \\ &= C^T A^i P_t C + C^T A^{i-1} k_t w_t \end{aligned} \tag{15}$$

where $P_t = E[X_t X_t^T]$ is the (dpxdp) zero-lag state correlation and is characterized by the Lyapunov equation

$$P_{t+1} = A P_{t+1} A^T + k_t w_t k_t^T \quad .$$

The right choice for the Kalman vector k_t and the correlation matrix w_t can make $\{r_i^t\}$ become time invariant.

Choose k_t and w_t such that

$$k_t w_t = -A P_t C + A Q_0 C + BW$$

and

$$w_t = r_0 - C^T P_t C \quad ,$$

where $\{r_t\}, B, Q_0$ are Markovian parameters defined in Section 2.2. Then,

$$\begin{aligned} r_i^t &= C^T A^i P_t C + C^T A^{i-1} [-A P_t C + A Q_0 C + BW] \\ &= C^T A^i Q_0 C + C^T A^{i-1} BW \\ &= r_i \quad . \end{aligned}$$

The nice result here is the following: if one is interested only in the second order properties of a stationary vector process, one can replace the time-invariant Markovian representation by a time varying innovations representation which gives the same mean value and correlation sequence and whose advantage is that the initial conditions are deterministically set at t=0.

Likelihood. Now we use the important property that the initial state in the innovations representation is set at zero. The innovations process is then

$$\hat{u}_t = y_t - \hat{y}_t\Big|_{t-1} : N(0, w_t) \quad .$$

Thus the vector of observation $\underline{y}^T$ will have the same distribution as the input vector $\underline{\hat{u}}^T$ with the different mean $(\hat{y}_t^T\Big|_{t-1}, \ldots, \hat{y}_{t+P-1}^T\Big|_{t-1}) = \underline{\hat{y}}^T$. Hence $\underline{y}$ is distributed as $N(\underline{\hat{y}}, \Omega_N)$, where Ω_N is the $d(N+1)xd(N+1)$ diagonal matrix $\Omega_N = \text{diag}(w_N, \ldots, w_0)$. As a result the likelihood function is expressed as a finite product of normal $N_{y_t}(\hat{y}_t\Big|_{t-1}, w_t)$ densities:

$$L(\underline{y}) = \prod_{t=0}^{N} N_{y_t}(\hat{y}_t\Big|_{t-1}, w_t)$$

The log likelihood of the observations is formulated in the favorable form,

$$\log(L(\underline{y})) = -\frac{N+1}{2}\log(2\pi) - \frac{1}{2}\sum_{t=0}^{N}\log|w_t| - \frac{1}{2}\sum_{t=0}^{N}\hat{u}_t^T w_t^{-1}\hat{u}_t \qquad (16)$$

with

$$\hat{u}_t = y_t - \hat{y}_t\Big|_{t-1}$$

In this form, the log likelihood function depends only on the innovations values $\hat{u}_t$ and variance values w_t. Both $\hat{u}_t$ and w_t are non-stationary sequences that have a finite time dependence, and are given at each time t by the Kalman filter defined for the innovations representation. This means that we have now a recursive way of calculating the exact likelihood function of vector ARMA processes. The values so obtained are then fed into a non-linear optimization procedure of the Newton-Raphson type that provides an optimal set of parameters. This procedure is repeated till convergence is achieved, and maximum likelihood estimates of the ARMA parameters obtained. This supposes that the orders (p,q) of the ARMA has been determined.

2.4. Innovations representation and an important Chapman-Kolmogorov equation. The two previous paragraphs have been devoted to the derivation of the likelihood function of the process $\{y_t\}$. By comparing the likelihood expressions (11) and (16), obtained respectively for the Markovian and innovations representations, we see that a very important Chapman-Kolmogorov equation has been solved:

$$\prod_{t=0}^{N} N_y(\hat{y}_t\Big|_{t-1}, w_t) = \int_{\text{all } X_0} \prod_{t=0}^{N} N_{y_t}(y_t\Big|_{t-1}, W) N_{X_0}(0, Q_0)\, dX_0 \qquad (17)$$

$$\hat{y}_t\Big|_{t-1} = C^T \hat{X}_t \quad ; \quad \hat{y}_{0/-1} = 0$$

$$\hat{X}_t = A\,\hat{X}_{t-1} + k_t \hat{u}_{t-1} \quad ; \quad \hat{X}_0 \equiv 0$$

$$\hat{u}_t = y_t - \hat{y}_t\Big|_{t-1}$$

$$y_t\Big|_{t-1} = C^T X_t \qquad ; \qquad y_{0/-1} : N(C^T X_0,\ C^T Q_0 C)$$

$$X_t = A X_t + B u_{t-1} \ ; \qquad X_0 : N(0, Q_0)$$

$$u_t = y_t - y_t\Big|_{t-1}$$

2.5. Comments. The maximum likelihood identification procedure may be conducted in the following way. First start with an initial guess of the ARMA parameters and compute the corresponding correlation sequence. (In the scalar case, [Dugre/Beex/Scharf, 1980] proposes an algorithm.) Then run the Kalman filter to obtain the value of the exact likelihood function. This value is fed into a non-linear optimization procedure that updates the vectors of parameters.

One may want to speed up the computations involved in Kalman filtering. Appealing to the formulae (14), it is seen that there exists a close relationship between the time varying impulse response sequence $\{h_i^t\}$ and the vector Kalman gain. We shall use this feature extensively in the next section to derive a Fast Kalman Algorithm (Morf, Sidhu, and Kailath algorithm) that will avoid the solution of the Ricatti equation. This also leads to a fast lattice implementation.

3. FAST ALGORITHMS

Recall the expressions of the exact likelihood (3) and (16) obtained by the conventional method and the innovations method. It is clear that

$$\text{Log}|R| = \sum_{t=0}^{N} \log|w_t|$$

and

$$-\frac{1}{2}\underline{u}^T\underline{u} = -\frac{1}{2}\sum_{t=0}^{N}(y_t-\hat{y}_t\Big|_{t-1})^T w_t^{-1}(y_t-\hat{y}_t\Big|_{t-1}) \quad .$$

Thus the innovations representation solve the triangularization of the inverse correlation matrix to obtain the white vector

$$\underline{u} = R^{-1/2}\,\underline{y} \quad ,$$

or since $R^{-1/2}$ is invertible

$$\underline{y} = R^{1/2}\,\underline{u} \quad . \tag{18}$$

But the triangular or Choleski decomposition is easily obtained from the innovations representation (12). If we write the observation vector $\underline{y}^T = [y_0^T,\ldots,y_N^T]$ upside down, we obtain the upper triangular block matrix equation:

$$\begin{bmatrix} y_N \\ \cdot \\ \cdot \\ \cdot \\ y_0 \end{bmatrix} = \begin{bmatrix} I & h_1^{N-1} & h_2^{N-2} & \cdots & h_N^0 \\ & I & h_1^{N-2} & \cdots & h_{N-1}^0 \\ & & \ddots & & \vdots \\ & 0 & & I & h_1^0 \\ & & & & I \end{bmatrix} \begin{bmatrix} \hat{u}_N \\ \cdot \\ \cdot \\ \cdot \\ \hat{u}_0 \end{bmatrix}$$

$$= K_N \begin{bmatrix} \hat{u}_N \\ \cdot \\ \cdot \\ \cdot \\ \hat{u}_0 \end{bmatrix}$$

Here the upper triangular block matrix K_N is (N+1)d x (N+1)d dimensional. The block elements $\{h_t^i\}$ of the matrix are defined as the time varying impulse response,

$$h_i^t = C^T A^{i-1} k_t$$

$$h_0^t = I$$

The block correlation matrix R is computed as the expected value of the outer product of the vector of observations:

$$R = E\left\{ \begin{bmatrix} y_N \\ \cdot \\ \cdot \\ \cdot \\ y_0 \end{bmatrix} [y_N^T \ \ldots \ y_0^T] \right\} \quad .$$

Thus the ((N+1)d x (N+1)d) dimensional block correlation matrix is decomposed into the following triangular form:

$$R = K_N \quad E\left\{ \begin{bmatrix} \hat{u}_N \\ \vdots \\ \hat{u}_0 \end{bmatrix} [\hat{u}_N^T \ \ldots \ \hat{u}_0^T] \right\} K_N^T \tag{19}$$

That is

$$R = \begin{bmatrix} r_0 & r_1 & \cdots & r_N \\ r_{-1} & r_0 & & \\ & \ddots & & \\ & & r_0 & r_1 \\ r_{-N} & & r_{-1} & r_0 \end{bmatrix} = \begin{bmatrix} I & h_1^{N-1} & & h_N^0 \\ & \ddots & & h_{N-1}^0 \\ 0 & & \ddots & \\ & & I & h_1^0 \\ & & & I \end{bmatrix} \begin{bmatrix} w_N & & & \\ & \ddots & 0 & \\ 0 & & \ddots & \\ & & & w_0 \end{bmatrix} \begin{bmatrix} I & & & \\ & \ddots & & 0 \\ & & \ddots & \\ & 0 & & I \end{bmatrix} \begin{bmatrix} I & & & \\ & \ddots & & \\ & & \ddots & \\ (h_N^0)^T & & & I \end{bmatrix} \tag{20}$$

Thus we may write

$$R = K_N \, \Omega_N \, K_N^T = R^{1/2} R^{1/2^T}$$

with

$$R^{1/2} = K_N (\Omega_N)^{1/2} \quad .$$

where Ω_N is the diagonal block matrix $\Omega_N = \text{diag}(w_N \ \ldots \ w_0)$. The triangular matrix K_N is invertible. Therefore we may write (20) in the form

$$K_N^{-1} \, R = \Omega_N \, K_N^T \tag{21}$$

By analogy with the scalar case, these equations are called the multivariate normal equations. The inverse of the matrix K_N, is redefined as $\underline{a}_N$:

$$\underline{a}_N = K_N^{-1} = \begin{bmatrix} I & a_1^N & \cdots & a_N^N \\ & I & & a_{N-1}^{N-1} \\ & & \ddots & \\ & 0 & & I \end{bmatrix} \tag{22}$$

Now the analogy with the scalar case is much clearer. The block vector $[I\ a_1^N,\ldots,a_N^N]$ represents the vector of parameters of an increasing order vector autoregressive process fitted on the correlation matrix of the process. This problem is solved in the scalar case by the Levinson-Durbin algorithm [Levinson, 1947], [Durbin, 1960]. Equivalently, one might solve (21) directly for the time varying impulse response h_i^t using the impulse response algorithm [Leroux/Gueguen, 1977]. [Robinson and Wiggins, 1965] entended the Levinson algorithm to the multivariate case.

3.1. The Generalized Impulse Response Algorithm. Here we want to derive an algorithm to obtain directly the matrix impulse responses h_i^t that appear in the matrix K_N. It must not be forgotten that we are after a fast algorithm to calculate the block Kalman gain vector k_t. From relation (14), it is seen that the matrix impulse response sequence is given by,

$$h_i^t = C^T A^{i-1} k_t \quad ,$$

where $C^T = [I\ 0,\ldots,0]$ and A is the block companion matrix given in Section 2.2. This means that the block Kalman gain vector k_t is composed of the first p impulse response matrices;

$$k_t^T = [(h_t^1)^T,\ldots,(h_t^p)^T] \quad .$$

The generalized impulse response algorithm provides a recursive method for calculating the Kalman gain vector.

Let's write the forward and backward predictors for orders p=0 to N;

Forward Prediction

$$\begin{bmatrix} I & a_1^N & a_N^N \\ & I & a_{N-1}^{N-1} \\ & 0 & \ddots \\ & & I \end{bmatrix} \begin{bmatrix} r_{-N} & \cdots & r_{-1} \\ r_{-N-1} & r_{-2} & \\ & \vdots & \\ r_{-2N} & r_{-N-1} & \end{bmatrix} \begin{bmatrix} r_0 & r_1 & & r_N \\ r_{-1} & r_0 & & \\ & & \ddots & \\ r_{-N} & & & r_0 \end{bmatrix} = \begin{bmatrix} S_N^N & \cdots & S_p^N & S_1^N \\ & & & S_2^{N-1} \\ & & & \vdots \\ & & & S_N^0 \end{bmatrix} \begin{bmatrix} S_0^N & & & \\ S_1^{N-1} & S_0^{N-1} & & 0 \\ & & \ddots & \\ S_N^0 & S_{N-1}^0 & & S_0^0 \end{bmatrix} \qquad (23)$$

Here $S_i^k = R_k^{\varepsilon}(h_i^k)^T$ and to connect the algorithm to the solution of (21), one

should have in mind that $R_k^{\varepsilon}=w_k$. The order p is the order of the AR part of the ARMA (p,q).

Backward Prediction

$$\begin{bmatrix} b_N^N & b_{N-1}^N & I \\ b_{N-1}^{N-1} & I & \\ \vdots & \ddots & \\ I & & \end{bmatrix} \begin{bmatrix} r_{-N} & r_{-1} \\ & r_{-2} \\ & \\ r_{-2N} & r_{-N-1} \end{bmatrix} \begin{bmatrix} r_0 & r_1 & r_N \\ r_{-1} & r_0 & \\ \vdots & \ddots & \\ r_{-N} & & r_0 \end{bmatrix} = \begin{bmatrix} V_{-2N}^N & \cdots & V_{-N-p}^N & \cdots & V_{-N-1}^N \\ & & & & V_{-N}^{N-1} \\ & & & & \\ & & V_{-p}^0 & \cdots & V_{-1}^0 \end{bmatrix} \begin{bmatrix} 0 & & & V_0^N \\ & & V_0^{N-1} & V_1^{N-1} \\ & \iddots & & \\ V_0^0 & \cdots & \cdots & V_N^0 \end{bmatrix} \tag{24}$$

From the matrix relation (23), the following results hold:

$$\begin{aligned} S_i^N &= \sum_{\ell=0}^{N} a_\ell^N r_{-i-\ell} \qquad a_0^N = I \quad \text{for all } N \\ S_i^N - S_i^{N-1} &= \sum_{\ell=1}^{N-1} (a_\ell^N - a_\ell^{N-1}) r_{-i-\ell} + a_N^N r_{-i-N} \quad . \end{aligned} \tag{25}$$

But the expression for the difference $a_\ell^N - a_\ell^{N-1}$ is given by the [Robinson-Wiggins algorithm, 1965] and hence substituting this expression in (25) yields

$$\begin{aligned} S_i^N - S_i^{N-1} &= -\alpha_{N-1}[R_{N-1}^r]^{-1} \sum_{\ell=1}^{N} B_{N-\ell}^{N-1} r_{-i-\ell} \\ b_0^N &= I \quad \text{for all } N \quad . \end{aligned} \tag{26}$$

We recognize from (24) that the sum is exactly

$$V_{-i-N}^{N-1} = \sum_{\ell=1}^{N} b_{N-\ell}^{N-1} r_{-i-\ell} \quad .$$

Therefore the elements S_i^N satisfy the recursion

$$S_i^N = S_i^{N-1} - \alpha_{N-1}[R_{N-1}^r]^{-1} V_{-i-N}^{N-1} \qquad i > 0 \quad . \tag{27}$$

The same derivation is applied to V_{-i-N-1}^N and it is easily shown that:

$$V_{-N-i-1}^N = V_{-N-i-1}^{N-1} - \beta_{N-1}[R_{N-1}^{\varepsilon}]^{-1} S_i^{N-1} \qquad i > 0 \quad . \tag{28}$$

The equations given R^r_{N+1} and R^ε_{N+1} do not need to be changed and remain valid provided that we can get an expression for $\alpha_N = \beta^T_N$ that does not include either the forward or the backward coefficients. From the Robinson-Wiggins algorithm we know that

$$\begin{aligned} a^N_N &= \alpha_{N-1}[R^r_{N-1}]^{-1} \\ b^N_N &= -\beta_{N-1}[R^\varepsilon_{N-1}]^{-1} \end{aligned} \qquad (29)$$

From (22) and (23), one can extract the expression for b^N_N given here:

$$b^N_N = -\beta_{N-1}[R^\varepsilon_{N-1}]^{-1} = -V^{N-1}_{-N}[S^{N-1}_0]^{-1} \quad . \qquad (30)$$

The same procedure leads to a similar expression for a^N_N:

$$a^N_N = -\alpha_{N-1}[R^r_{N-1}]^{-1} = -S^{N-1}_{-N}[V^{N-1}_0]^{-1} \quad . \qquad (31)$$

The generalized impulse response algorithm is then summarized as follows:

$$\begin{aligned} S^N_i &= S^{N-1}_i - \alpha_{N-1}[R^r_{N-1}]^{-1} V^{N-1}_{-i-N} \qquad \text{for all } i \\ V^N_i &= V^{N-1}_i - \beta_{N-1}[R^\varepsilon_{N-1}]^{-1} S^{N-1}_{-i-N} \qquad \text{for all } i \\ R^r_N &= R^r_{N-1} - \beta_N[R^\varepsilon_N]^{-1} \alpha_N \\ R^\varepsilon_N &= R^\varepsilon_{N-1} - \alpha_N[R^r_N]^{-1} \beta_N \end{aligned} \qquad (32)$$

and

$$\begin{aligned} \alpha_{N-1}[R^r_{N-1}]^{-1} &= S^{N-1}_{-N}[V^{N-1}_0]^{-1} \\ \beta_{N-1}[R^\varepsilon_{N-1}]^{-1} &= V^{N-1}_{-N}[S^{N-1}_0]^{-1} \end{aligned}$$

Initial Conditions. The initial conditions are straightforward and given by

$$\begin{aligned} [V^0_0,\ldots,V^0_N] &= [r_0,\ldots,r_N] \\ [S^0_N,\ldots,S^0_0] &= [r_{-N},\ldots,r_0] \quad . \end{aligned}$$

In the scalar case it reduces to the impulse response algorithm proposed by [Leroux/Gueguen, 1977]. By analogy, the terms $\alpha_{N-1}[R^r_{N-1}]^{-1}$ and $\beta_{N-1}[R^\varepsilon_{N-1}]^{-1}$ will be called multivariate forward and backward reflection coefficients (or parcor coefficients).

Remarks. This algorithm, although a recursive way of computing the impulse response of the increasing order vector AR processes, does not improve the number of computations a great deal. In fact, if one wants to obtain the Kalman vector $K_N^T = [(h_1^N)^T,\ldots,(h_p^N)^T]$, one has to start the algorithm with the knowledge of the matrix correlation sequence for $-n-p-1 \leq i \leq p$. This interval depends on N.

So far the generalized impulse response algorithm has been derived without using the information we have of the internal structure for the process.

3.2. Generalized Impulse Response Algorithm for ARMA Processes: The Morf, Sidhu and Kailath (MSK) Algorithm. We show here that the generalized impulse response algorithm, modified to take account of the ARMA structure, is identical to the [Morf, Sidhu and Kailath, 1974] algorithm derived from the Chandrasekar type equations. As the process $\{y_t\}$ is ARMA, it behaves like an AR on its tail. That is, for $N \geq 0$, the matrix correlation sequence $\{r_t\}$ satisfies the AR(p) recursion,

$$r_{N+1+p} + a_1 r_{N+p} + \ldots + a_p r_{N+1} = 0 \qquad \text{for all } N \geq 0. \tag{33}$$

Here the $\{a_i\}$ are the true parameters of the autoregression. After transposition, equation (33) becomes,

$$r_{-N-1-p} = [r_{-N-1},\ldots,r_{-N-p}] \begin{bmatrix} -a_p^T \\ \vdots \\ -a_1^T \end{bmatrix} \tag{34}$$

If we write this equation for r_{-p-1} to r_{-N-1-p}, then the following matrix equation is true:

$$\begin{bmatrix} r_{-p-1} \\ \cdot \\ \cdot \\ \cdot \\ r_{-N-p-1} \end{bmatrix} = \begin{bmatrix} r_{-1} & r_{-2} & & r_{-1} \\ r_{-2} & & & r_{-p-1} \\ & & & \\ r_{-N-1} & & & r_{-N-p} \end{bmatrix} \begin{bmatrix} -a_p^T \\ \cdot \\ \cdot \\ \cdot \\ -a_1^T \end{bmatrix} \tag{35}$$

Premultiply by the row block vector of backward prediction coefficients and obtain:

$$[b_N^N, \ldots, b_1^N \quad I] \begin{bmatrix} r_{-p-1} \\ r_{-p-2} \\ \vdots \\ r_{-N-p-1} \end{bmatrix} = V_{-N-p-1}^N \quad . \tag{36}$$

Therefore the internal structure imposed by the ARMA nature of the process is characterized by the relation

$$V_{-N-p-1}^N = [V_{-N-1}^N \; V_{-N-2}^N \cdots V_{-N-p}^N] \begin{bmatrix} -a_p^T \\ -a_{p-1}^T \\ \cdot \\ \cdot \\ -a_1^T \end{bmatrix} \quad . \tag{37}$$

This indicates that we need only to compute p values V_{-N-1}^N to V_{-N-p}^N. The initial conditions are now given by $\{r_i\}_{-p}^{+p}$, independent of N. This observation summarizes the essence of the Fast Kalman algorithm and underlies the MSK algorithm. From (37), the vector $[V_{-N-2}^N \cdots V_{-N-p-1}^N]$ is given by the linear transformation,

$$[V_{-N-2}^N \cdots V_{-N-p-1}^N] = [V_{-N-1}^N \cdots V_{-N-p}^N] A^T \tag{38}$$

where A is the state companion matrix. The recursions giving S_i^N and V_i^N in the generalized impulse response algorithm (32) are written for i=1, ...,p:

$$\begin{aligned} S_i^{N+1} &= S_i^N - \alpha_N [R_N^r]^{-1} V_{-N-i-1}^N \qquad i=1,\ldots,p \\ V_{-N-i-1}^{N+1} &= V_{-N-i-1}^N - \beta_N [R_N^\varepsilon]^{-1} S_i^N \end{aligned} \tag{39}$$

Define now the following (dxdp) dimensional block vectors

$$\underline{K}_t = \begin{bmatrix} (S_1^t)^T \\ \vdots \\ (S_p^t)^T \end{bmatrix} \quad \text{and} \quad \underline{V}_t = \begin{bmatrix} (V_{-t-1}^t)^T \\ \vdots \\ (V_{-t-p}^t)^T \end{bmatrix}$$

Then the set of equations (39) is condensed into the vector form

$$\begin{aligned} \underline{K}_{N+1}^T &= \underline{K}_N^T - \alpha_N [R_N^r]^{-1} \underline{V}_N^T A^T \\ \underline{V}_{N+1}^T &= \underline{V}_N^T A^T - \beta_N [R_N^\varepsilon]^{-1} \underline{K}_N^T \quad . \end{aligned} \tag{40}$$

It is also noted that $R_N^\varepsilon = S_0^N$, and $R_N^r = V_0^N$. These identities, together with the expressions for $\alpha_N[R_N^r]^{-1}$ and $\beta_N[R_N^\varepsilon]^{-1}$ given in (32), prove that:

$$\begin{aligned} [V_{-N}^{N-1}]^T &= C^T \underline{V}_N \qquad C^T = [I\ 0,\ldots,0] \\ \alpha_N^T &= \underline{V}_N^T C = \beta_N \end{aligned} \tag{41}$$

Finally, for purposes of easy comparison with the MSK algorithm, define the matrix M_k as the inverse of the matrix of mean square backward prediction errors:

$$M_k = [R_k^r]^{-1}$$

Then the generalized impulse response algorithm for a vector ARMA process is summarized in these equations:

$$\begin{aligned} \underline{K}_{N+1} &= \underline{K}_N - A \underline{V}_N M_N^T \underline{V}_N^T C \\ \underline{V}_{N+1} &= A \underline{V}_N - \underline{K}_N [R_N^\varepsilon]^{-T} C^T \underline{V}_N \\ [R_{N+1}^\varepsilon]^T &= [R_N^\varepsilon]^T + C^T \underline{V}_N [M_N]^T \underline{V}_N^T C \\ [M_{N+1}]^T &= [M_N]^T - [M_N]^T \underline{V}_N^T C\, [R_N^\varepsilon]^{-1} C^T \underline{V}_N [M_N]^T \quad . \end{aligned} \tag{42}$$

This algorithm is readily recognized as one form of the MSK algorithm [Morf/Sidhu/Kailath, 1974], [Friedlander et al. 1978], applied on the innovations state space model. To complete the identification the reader should note that the matrix C^T and the block vector $\underline{V}_N$ in (42) correspond

respectively to the matrix H and the block vector $\underline{Y}_N$ in the MSK algorithm. The number of computations is significantly reduced compared to the direct solution of the Riccati equation or to the generalized impulse response algorithm. It should be clear that a number of operations can be performed in parallel.

Remarks. Recall that we are after an algorithm to compute the block Kalman gain vector in order to calculate the exact likelihood function. The vector $\underline{K}_t$ introduced in (40) is defined as

$$\underline{K}_t = k_t \, \Omega_t$$

where k_t is the block Kalman gain vector we want and Ω_t is the block diagonal correlation matrix of the innovations process, $\Omega_t = \text{diag}(w_t, \ldots, w_0)$.

The same derivation can be worked out in the scalar case starting from the scalar impulse response algorithm. It leads to the scalar version of the MSK algorithm used by [Pearlman, 1980].

The generalized impulse response algorithm or the fast Kalman algorithm are also fast algorithms to generate the generalized reflexion coefficients $\alpha_{N-1}[R^r_{N-1}]^{-1}$ and $\beta_{N-1}[R^c_{N-1}]^{-1}$. Hence the method can be implemented as a fast lattice as well.

4. CONCLUSIONS

In this paper we have derived a recursive procedure to compute the exact likelihood for vector ARMA processes. The key to the method was the "innovations" representation of the process that allowed the use of Kalman filtering techniques. The Kalman vector gain was shown to be composed of time varying impulse response values of the process. This motivated the derivation of the generalized impulse response algorithm. Finally, introducing the knowledge of the internal structure of the ARMA process, we showed that the generalized impulse response algorithm was identical to the MSK or fast Kalman algorithm derived from the Chandrasekar type equations.

REFERENCES

G.E.P. Box and G. M. Jenkins, "Time Series Analysis: Forecasting and Control," Holden-Day, San Francisco, CA, 1976.

P. Newbold, "The Exact Likelihood Function of a Mixed Autoregressive Moving Average Process, Biometrika, 61, 1974.

M. M. Ali, "Analysis of Autoregressive Moving Average Models: Estimation and Prediction," Biometrika, 64, 1977.

D. R. Osborn, "Maximum Likelihood Estimation of Moving Average Processes," Annals of Economic and Social Measurements, 511, 1976.

M. S. Phadke and G. Kedem, "Computation of the Exact Likelihood Function of Multivariate Moving Average Models," Biometrica, 65, 1978.

G. F. Ansley, "An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process," Biometrika, 66, 1979.

H. Akaike, "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models," Biometrika, 60, 1973.

T. W. Anderson, "Maximum Likelihood Estimation for Vector Autoregressive Moving Average Models," Tech. Rep. No. 35, Dept. Stat., Stanford University, 1978.

R. L. Kashyap, "Maximum Likelihood Identification of Stochastic Linear Systems," IEEE Trans. on Aut. Cont., AC-15, February 1970.

F. Arhabi, "Maximum Likelihood Estimation of the Covariances of the Vector Moving Average Models in the Time and Frequency Domains," Tech. Rep. No. 36, Dept. Stat., Stanford University, 1978.

F. Arhabi, "Maximum Likelihood Estimation of the Autoregressive Coefficients and Moving Average Covariances of Vector Autoregressive Moving Average Models," Tech. Rep., Dept. Stat., Stanford University, 1979.

S. A. Tretter and K. Steiglitz, "Power Spectrum Identification in Terms of Rational Models," IEEE Trans. Aut. Cont., AC-12, 1967.

G. Tunnicliffe Wilson, "The Estimation of Parameters in Multivariate Time Series Models," Journal of the Royal Statistical Society, Vol. 35, 1973.

W. Dunsmuir and E. J. Hannan, "Vector Linear Time Series Models," Advances in Applied Probability, Vol. 8, 1976.

P. Shar , "On the Inverse of the Covariance Matrix for an Autoregressive Moving erage Process," Biometrika, 60, 1973.

R. ' .lbraith and I. I. Galbraith, "On the Inverse of Some Patterned Matrices Arising in the Theory of Stationary Time Series," Journal Applied Probability, 11, 1974.

F. C. Schweppe, "Evaluation of Likelihood Functions of Gaussian Signals," IEEE Trans. Inf. Th., IT-11, 1965.

A. C. Harvey, G.D.A. Phillips, "Maximum Likelihood Estimation of Regression Models with Autoregressive Moving Average Disturbances," Biometrika, 66, 1979.

R. H. Jones, "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations," Technometrics, August 1980.

H. Akaike, "Markovian Representation of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes," Ann. Inst. Statist. Math., 26, 1974.

C. Gueguen and L. L. Scharf, "Exact Maximum Likelihood Identification of ARMA Models; A Signal Processing Perspective," Proc. EUSIPCO, September 1980, Lausanne.

W. A. Wolovich, "Linear Multivariate Systems," Springer-Verlag, 1974.

J. P. Dugré, A. A. Beex, L. L. Scharf, "Generating Covariance Sequences and the Calculation of Quantization and Rounding Error Variances in Digital Filters," IEEE Trans. on ASSP, ASSP-28, Feb. 1980.

J. G. Pearlman, "An Algorithm for the Exact Likelihood of a High Order ARMA Process," Biometrika, 67, 1, 1980.

M. Morf, G. S. Sidhu and T. Kailath, "Some New Algorithms for Recursive Estimation on Constant Linear Discrete Time System," IEEE Trans. on Aut. Cont., AC-19, August 1974.

B. Friedlander, T. Kailath, M. Morf and L. Ljung, "Extended Levinson and Chandrasekar Equations for General Discrete Time Linear Estimation Problems," IEEE Trans. on Aut. Cont., AC-23, August 1978.

R. A. Wiggins and E. A. Robinson, "Recursive Solution to the Multivariate Filtering Problem," J. Geophys, Res., Vol. 70, April 1965.

N. Levinson, "The Wiener RMS (root-mean-square) Error Criterion in Filter Design and Prediction," J. Math. Phys., Vol. 25, January 1947.

J. Durbin, "The Fitting of Time Series Models," Rev. Int. Stat. Inst., Vol. 28, 1980.

J. Leroux and C. Gueguen, "A Fixed Point Computation of the Partial Correlation Coefficients," Trans. on ASSP, Vol. 25, June 1977.

B.D.D. Anderson and J. B. Moore, "Optimal Filtering," Prectice Hall, Inf. Syst. Series, 1978.

TOPICS IN PARAMETRIC SPECTRUM ANALYSIS: THE ARMA CASE

J. P. Dugre
C. J. Gueguen
N. Moreau
L. L. Scharf

Presented at:

JOA-UAG Conference on Spectral Analysis and Its
Application to Underwater Acoustics, London
April 1982

TOPICS IN PARAMETRIC SPECTRUM ANALYSIS: THE ARMA CASE

J. P. Dugre* C. J. Gueguen** N. Moreau** L. L. Scharf*

*Electrical Engineering Dept. University of Rhode Island Kingston, RI 02881

**Ecole Nationale des Telecommunications, 46 rue Barrault 75634 Paris, France

ABSTRACT

A general framework for deriving and interpreting analysis and synthesis spectra of the autoregressive (AR) and moving average (MA) type is reviewed. Linear transformations of finite data records provide for a unified treatment of spectra as different as the Bartlett (BA), the maximum likelihood method (MLM) and the maximum entropy method (MEM) spectra. We then generalize these ideas to ARMA linear transformations. An ARMA type spectrum is obtained and related to other ARMA spectra. A parameterization scheme is proposed.

INTRODUCTION

In many applications of signal processing such as sonar, radar, speech, and communications, only short segments of the processes are available. This situation arises due either to inherent nonstationarities that force segmentation or to the short interval over which the signal can be observed. Hence the problem of estimating spectral density functions from a finite set of time series observations is a crucial step in the modeling of underlying data. Current research activity is primarily centered on "high resolution" or parametric methods of spectrum analysis (cf ex: (1)-(4)). Recently a general framework for deriving and interpreting analysis and synthesis spectra of the autoregressive (AR) and moving average (MA) type has been introduced (5)(6).

In most techniques, stationary time series are modeled as the output of time invariant linear systems (filters) driven by stationary white noise sequences. When dealing with finite length data records the problem of initial conditions arises and is generally avoided by making some assumptions on the data outside

the interval of observation.

Alternatively, the proper initialization may be obtained by identifying time varying models where the initial conditions manifest themselves as time variations. In this approach, the vector of data is viewed as the result of a linear transformation applied to a vector of uncorrelated values. This formulation leads to a unified presentation and interesting interpretations of spectral estimates such as the Bartlett (BA), maximum likelihood method (MLM) and maximum entropy method (MEM) spectra (5),(6). Depending on the type of data being analyzed we are able to improve the performance of the BA and MLM spectra using a parameterization procedure. It appears that these procedures provide a method of classifying data as AR, MA or ARMA. These results are briefly reviewed in the first part of this paper.

We generalize these concepts to the definition of an ARMA linear transformation on finite length data records using results from (7) and the statistics literature (8), (9). This leads us to the derivation of an ARMA spectral estimate analogous to the MLM spectrum in the AR case or the BA spectrum in the MA case. Its relationship to "high resolution" ARMA spectra is also discussed. As in the AR and MA case, it is speculated that an appropriate parameterization procedure would improve the performance of such a spectral estimate, at least for ARMA type data.

AR AND MA LINEAR TRANSFORMATIONS

Let $Y_t = (y_o \ldots\ldots y_{t-1})^T$ denote a t-sample snapshot of the real, zero mean, wide sense stationary process (y_t) with correlation sequence (r_t). Y_t has a symmetric and Toeplitz correlation matrix with first row $(r_o, \ldots\ldots r_{t-1})$. Let $U_t = (u_o \ldots\ldots u_{t-1})^T$ be a white vector with uncorrelated entries such that $E(U_t) = 0$ and $E(U_t U_t^T) = I_t$ (txt identity). The AR and MA linear transformations are defined respectively by the matrix equations

$$A_t Y_t = U_t$$

$$Y_t = H_t U_t$$

where the lower triangular matrices A_t and H_t are obtained by a Gram-Schmidt orthogonalization procedure. These matrices can be computed from the correlation matrix or its inverse since the following Cholesky decompositions hold and are unique:

$$R_t = H_t H_t^T$$

$$R_t^{-1} = A_t^T A_t$$

$$A_t H_t = I_t$$

Generalizing the linear system concept of frequency response, we define the "frequency response" of the AR and MA linear transformations in a natural way as[1]

$$A_t(\omega) = A_t\, C_t(\omega)$$

$$H_t(\omega) = C_t^H(\omega)\, H_t$$

with

$$C_t^H(\omega) = (1\ e^{j\omega}\ \ldots.e^{j\omega\,(t-1)})$$

AR Case: As A_t is a whitening transformation, we associate with the AR complex frequency response the spectrum

$$R_t(\omega) = \frac{1}{|\, A_t(\omega)\,|^2} = \frac{1}{C_t^H(\omega)\, R_t^{-1}\, C_t(\omega)}$$

This spectrum is recognized as the maximum likelihood method spectrum. In the case of AR(p) data, the AR linear transformation A_t is $A_{t,p}$:

[1] Superscript (H)T denotes (Hermitian) transpose.

$$A_{t,p} = \begin{bmatrix} a_o^o & & & 0 \\ a_1^1 & a_o^1 & & \\ & & & \\ a_p^p & \cdots & a_o^p & \\ & \ddots & & \ddots \\ 0 & & a_p^p & \cdots & a_o^p \end{bmatrix}$$

In general the AR transformation matrix A_t can be approximated by $A_{t,p}$ where the matrix is completed by repetition of the order -p row. The associated spectrum is called a parameterized MLM spectrum and has been shown to yield good spectral estimates in the case of AR type data. (See (5), (6), and Figure 1 of this paper).

MA case: The spectrum associated with the MA linear transformation is

$$R_t(\omega) = |H_t(\omega)|^2 = C_t^H(\omega)\, H_t H_t^T C_t(\omega).$$

This spectrum is a scaled version of the conventional or BA spectrum as written in (10). If the data were MA(q), then the MA linear transformation matrix $H_{t,q}$ would be banded even though the rows would remain time varying:

$$H_t = \begin{bmatrix} h_o^o & & & 0 \\ & \ddots & & \\ h_q^q & & h_o^q & \\ & \ddots & & \ddots \\ h_t^t & & h_q^t & & h_o^t \end{bmatrix} \begin{bmatrix} h_o^o & & & 0 \\ & \ddots & & \\ h_q^q & & h_o^q & \\ & \ddots & & \ddots \\ 0 & & h_q^q & & h_o^q \end{bmatrix} = \hat{H}_{t,q}$$

A parameterization scheme $\hat{H}_{t,q}$ insuring non-negative spectra is proposed and associated with a parameterized BA spectrum in (6). (See Figure 2).

ARMA LINEAR TRANSFORMATIONS

Suppose the underlying process (y_t) is ARMA (p,q):

$$\sum_{j=o}^{p} a_j \, y_{t-j} = \sum_{j=o}^{q} b_j \, \varepsilon_{t-j}$$

Here $a_o = b_o = 1$ and ε_t is a white noise process. The snapshot y_t has a correlation matrix R_Y that has the LU decomposition

$$R_Y = H_t \, H_t^T \quad .$$

We define the AR part of (y_t) as the AR process (z_t) with coefficients $(a_o \ldots, a_p)$. A snapshot of the process Z_t satisfies the AR linear transformation

$$A_{t,p} \, Z_t = U_t$$

where U_t is a vector of uncorrelated data and $A_{t,p}$ is given by the Cholesky decomposition

$$R_Z^{-1} = A_{t,p}^T \, A_{t,p}$$

Recall $A_{t,p}$ has the banded structure illustrated earlier. Applying this AR linear transformation to the snapshot Y_t, produces a finite length vector X_t whose correlation matrix R_X has the following structure (7), (8), and (9):

$$R_X = \begin{bmatrix} & & & & 0 \\ & & & \iddots & \\ & & \iddots & & \\ & \iddots & & & \\ 0 & & & & \end{bmatrix} \quad \begin{matrix} \\ \\ \\ \\ \text{p} \end{matrix}$$

$$A_{t,p} \, Y_t = X_t$$

$$A_{t,p} R_Y A_{t,p}^T = R_X$$

The matrix R_X is banded Toeplitz except for the upper left (p x p) corner. Thus R_X is almost the (t x t) correlation matrix of a pure MA(q). However, R_X has an L U Cholesky decomposition as follows:

$$R_X = B_X \, B_X^T$$

B_X has the particular lower triangular structure (8)(9),

$$B_{t,q} = \begin{bmatrix} b_o^o & & & & \\ & & & & 0 \\ b_p^p & & b_o^p & & \\ 0 & b_q^{p+1} & & b_o^{p+1} & \\ & 0 & & & \\ & & b_q^t & & b_o^t \end{bmatrix}$$

This decomposition corresponds to an MA linear transformation applied to E_t to produce X_t:

$$X_t = B_{t,q} E_t \quad .$$

Here E_t is a vector of uncorrelated data. As a result the ARMA linear transformation is summarized by the matrix relations

$$A_{t,p} Y_t = X_t = B_{t,q} E_t$$

$$A_{t,p} H_t = B_{t,q} \quad .$$

It is not hard to show that H_t is an impulse response matrix determined by the Kalman gain (7). This fully characterizes $B_{t,q}$ and sharpens the result of Newton and Pagano (8), (9).

It should be noted that $B_{t,q}$ is exactly a banded matrix for $q \geq p-1$. The matrix $A_{t,p}$ is easily constructed from $(a_o \ldots a_p)$ by running the Levinson/Durbin algorithm for decreasing j,

$$a_\ell^k = a_\ell \qquad k \geq p, \; o \leq \ell \leq k$$

$$a_i^j = (a_i^{j+1} - a_{j+1}^{j+1} a_{j+1-i}^{j+1})(1-(a_{j+1}^{j+1})^2)^{-1} \qquad j < p, \quad o \leq i \leq j$$

It is also shown in (8) that the last row of $B_{t,q}$ converges to the vector of true MA coefficients as t goes to infinity:

$$\lim_{t \to \infty} b_i^t = b_i \qquad o \leq i \leq q$$

ARMA Spectrum: If $S_X(\omega)$ and $S_Y(\omega)$ denote the spectra corresponding to X_t and Y_t, a plausible definition of $S_Y(\omega)$ is

$$S_Y(\omega) = \frac{S_X(\omega)}{C^H(\omega)A^T_{t,p}A_{t,p}C(\omega)}$$

$$S_X(\omega) = C^H(\omega)B_{t,q}B^T_{t,q}C(\omega)$$

Hence an ARMA spectrum has the following form:

$$S_Y(\omega) = \frac{C^H(\omega)B_{t,q}B^T_{t,q}C(\omega)}{C^H(\omega)A^T_{t,p}A_{t,p}C(\omega)} = \frac{C^H R_X C}{C^H R_Z^{-1} C}$$

This spectrum reduces to the MLM (BA) spectrum in the event of AR (MA) data.

ARMA Spectrum Analysis:

The method is based on the initial computation of the AR coefficients. For $t > p+q$, these coefficients are computed using the lower left corner of zeros in R_X

$$\begin{bmatrix} & & 0 \\ a_p & a_o & A_{t,p} \\ 0 & & \\ & a_p & a_q \end{bmatrix} \begin{bmatrix} r_y(o) & & r_y(t) \\ & & \\ r_y(t) & & r_y(o) \end{bmatrix} A^T_{t,p} = \begin{bmatrix} & & 0 \\ & & \\ 0 & & \end{bmatrix}$$

In fact this corresponds to solving the normal equations on the tail:

$$\sum_{k=1}^{p} a_k \, r_y(m-k) = -r_y(m). \quad m \geq q+1$$

The method can be summarized in the following steps.

- Estimate the correlation matrix R_Y, using the covariance method of linear prediction, for example.

- Solve for $(a_o \ldots a_p)$, using the normal equations on the tail.

- Decompose $R_Y = H_t H_t^T$ using a fast Choslesky decomposition routine

- Compute $B_{t,q} = A_{t,p} H_t$

- Compute $S_Y(\omega) = C^H B_{t,q} B_{t,q}^T C / C^H A_{t,p}^T A_{t,p} C$

In the limiting case,

$$\lim_{t \to \infty} S_Y(\omega) = \frac{C^H \underline{b}_q \underline{b}_q^T C}{C^H \underline{a}_p \underline{a}_p^T C}$$

where $\underline{b}_q = (b_o \ldots . b_q)^T$ and $\underline{a}_p = (a_o \ldots . a_p)^T$.

Relationship With The "High Resolution" ARMA Spectrum: The "high resolution" ARMA spectrum (4) can be written using our notation as

$$S(\omega) = \frac{S_\varepsilon(\omega)}{C^H \underline{a}_p \underline{a}_p^T C}$$

where $\underline{a}_p$ is the vector of AR coefficients and is computed by solving the normal equations on the tail, corresponding in fact to using the lower left corner of zeros in R_X. $S_\varepsilon(\omega)$ is the smoothed periodogram of the filtered process (ε_k) $k = 1 \ldots . t$:

$$\varepsilon_k = y_k + \sum_{i=1}^{p} a_i y_{k-i}$$

This time invariant filter can be expected to whiten the data less than the linear transformation $A_{t,p}$ and hence yield a process ε_k far from a pure MA(q).

The spectrum proposed here is consistent with both the MLM and the BA spectrum. Furthermore it is fairly flexible since one can apply a parameterization procedure on the AR part or on the MA part or both. Different sizes can be chosen to extend the matrices $A_{t,p}$ and $B_{t,q}$ depending on the relative importance of the poles and zeros in the spectrum. This should give better control of the compromise between resolution and smoothness and provide a tool for analysis of spectra of the ARMA type. For the moment, these results are conjecture based on the experience gained for AR and MA linear transformations.

REFERENCES

(1) J. P. Burg, "Maximum Entropy Spectral Analysis," Ph.D. Dissertation, Stanford University, 1975.

(2) T. J. Ulrych, T. N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," Rev. Geoph. Space. phys, Vol 13, Feb. 1973.

(3) S. Kay, "A New ARMA Spectral Estimator," IEEE Trans. on ASSP, ASSP.28, Oct. 1980.

(4) J. A. Cadzow, "High Performance Spectral Estimation. A New ARMA Method," IEEE Trans. on ASSP, ASSP-28, Oct. 1980.

(5) L. L. Scharf, C. J. Gueguen, J. P. Dugre, "Parametric Spectrum Modelling: A Signal Processing Perspective," First IEEE Workshop on Spectrum Analysis, Hamilton, Aug. 1981.

(6) L. L. Scharf, C. J. Gueguen, J. P. Dugre, N. Moreau, "Linear Transformations and Parametric Spectrum Analysis," IEEE Conf. on ASSP, Paris, May 1982.

(7) C. Gueguen and L. L. Scharf, "Exact Maximum Likelihood Identification of ARMA Models: A Signal Processing Perspective," Signal Processing: Theories and Applications, M. Kant and F. de Coulon (Eds.), North Holland, EURASIP (1980), Lausanne.

(8) M. J. Newton, M. Pagano, "The Finite Memory Prediction of Covariance Stationary Time Series," Technical Rep. n°21, Texas AM University, April 1981.

(9) M. J. Newton, M. Pagano, "Minimum Mean Square Error Prediction of Autoregressive Moving Average Time Series," Technical Rep. no.24, Texas AM University, Inst. of Stat., April 1981.

(10) R. T. Lacoss, "Data Adaptive Spectral Analysis Methods," Geophysics, Vol. 26, 4, Aug. 1971.

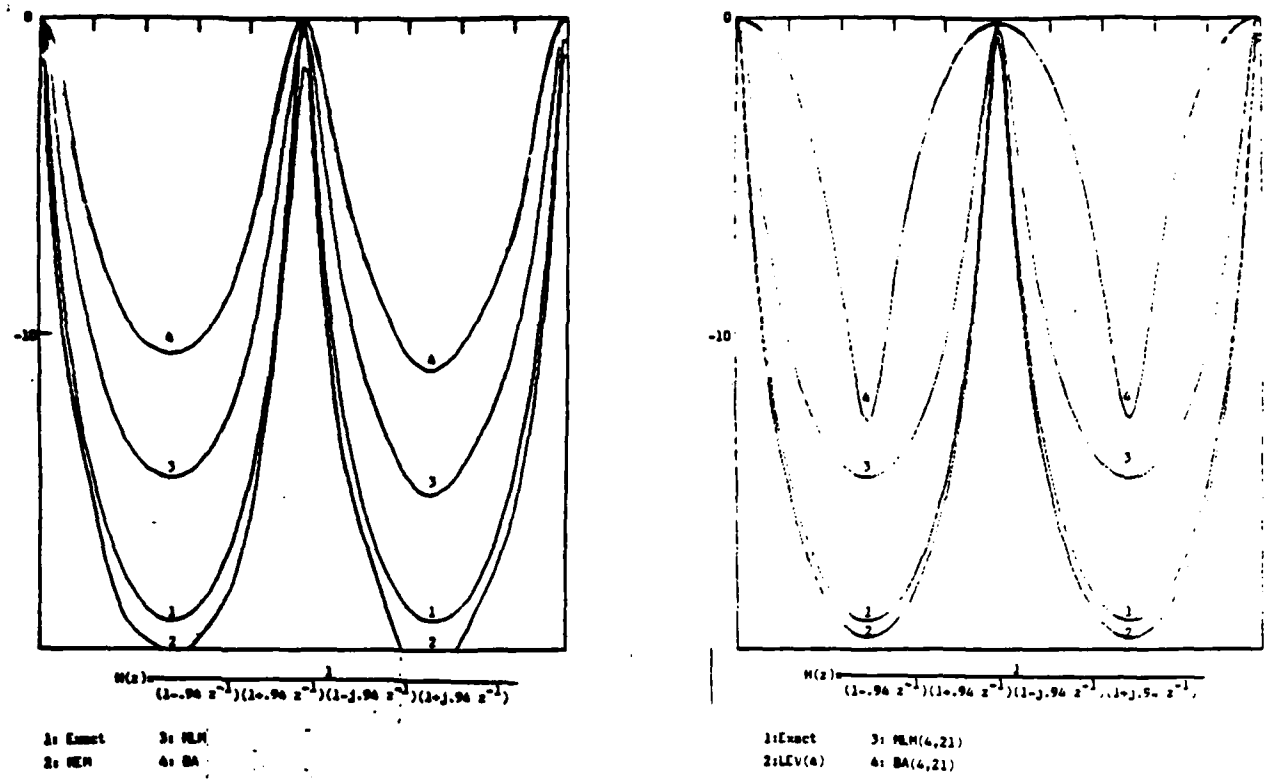


Fig.1. Influence of Parameterization;AR case

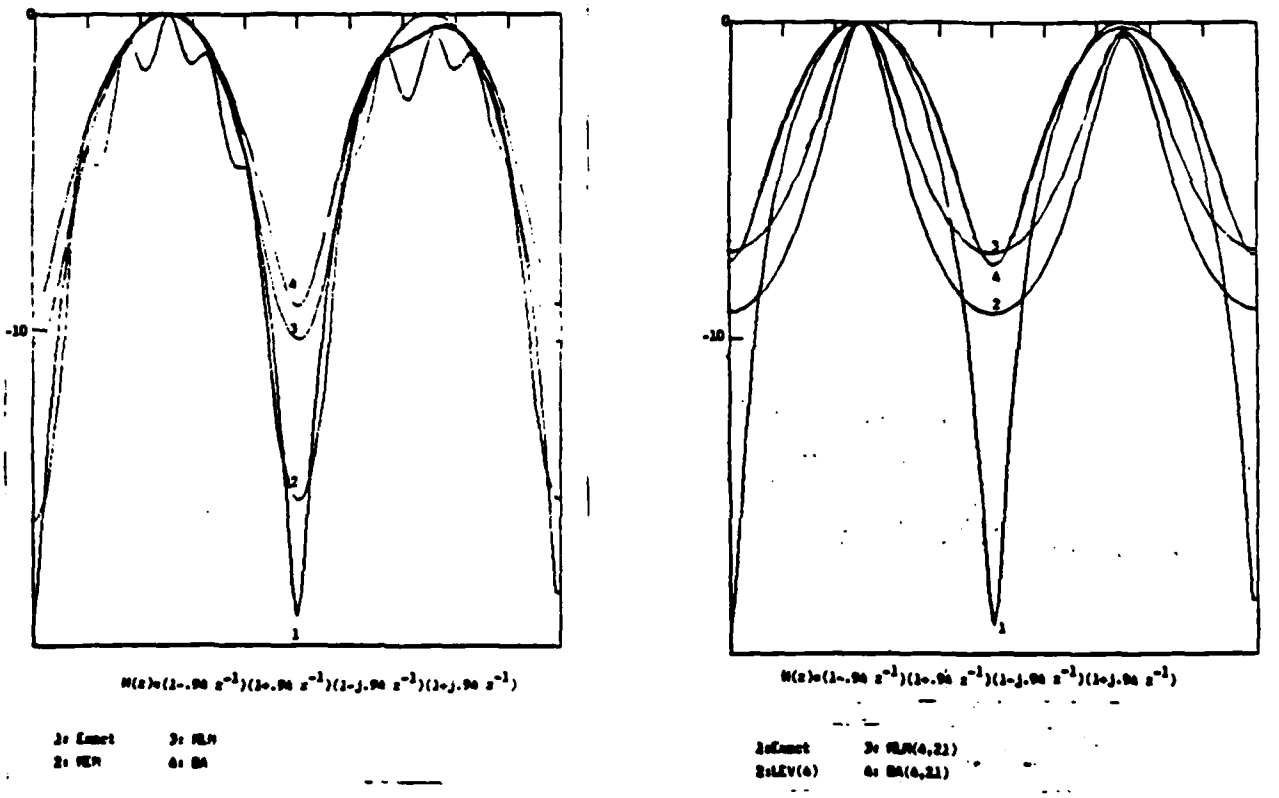


Fig.2. Influence of Parameterization;MA case

LINEAR TRANSFORMATIONS AND PARAMETRIC SPECTRUM ANALYSIS

L. L. Scharf
C. J. Gueguen
J. P. Dugre
N. Moreau

Presented at:

IEEE Conference on Acoustics, Speech, and Signal Processing, Paris
May 1982

LINEAR TRANSFORMATIONS AND PARAMETRIC SPECTRUM ANALYSIS

L. L. Scharf* C. J. Gueguen** J. P. Dugre* N. Moreau**

*Electrical Engineering Department
University of Rhode Island
Kingston, RI 02881

**Ecole Nationale Superieure des
Telecommunications
46 rue Barrault, Paris 75634

ABSTRACT

A general framework for deriving and interpreting analysis and synthesis spectra of the autoregressive (AR) and moving average (MA) type is presented. Investigation of AR linear transformations of finite dimensional data records yields a set of intermediate AR techniques associated with approximation of the inverse correlation matrix R^{-1}. The corresponding spectrum we call a parameterized maximum likelihood method (pMLM) spectrum. Investigation of MA linear transformations yields a set of intermediate MA techniques associated with approximation of the correlation matrix R. The corresponding spectrum we call a parameterized Bartlett spectrum (pBA).

Simulations on synthetic AR, MA and ARMA data sets illustrate the techniques and lead to interesting remarks concerning the use of parameterizations of R and R^{-1} to differentiate between data sets of AR and MA type.

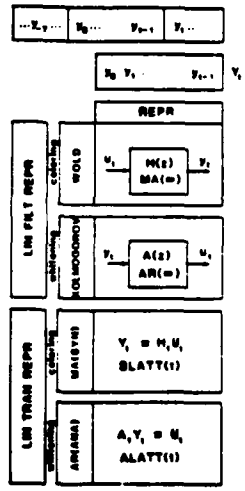
INTRODUCTION

Modern spectrum estimation is primarily concerned with the identification of parametric models that represent an underlying random process. In most techniques, the concept of wide sense stationarity seems to underly the very notion of a spectrum. A stationary time series is modeled as the output of a time invariant linear system driven by a stationary white noise process. The difficulty with stationary models is that initial conditions manifest themselves as nuisance parameters. The problem of the proper initialization can be solved by identifying non stationary models where the inital conditions are absorbed naturally into the theory.

Pursuing ideas developed in [1], we present a general framework for deriving and interpreting analysis and synthesis spectra of the autoregressive (AR) and moving average (MA) type. Investigation of AR linear transformations of finite dimensional data records yields a set of intermediate techniques (parameterized likelihood) [1], [2], associated with the maximum likelihood method (MLM) [3] and maximum entropy method (MEM) [4], [5], [6], of spectrum analysis. Investigation of MA linear transformations yields intermediate MA techniques (parameterized Bartlett) associated with an approximation of the correlation matrix and and the conventional Bartlett (BA) spectrum [7].

Simulations on synthetic data sets lead to interesting conclusions: the parameterizations of the correlation matrix R and its inverse R^{-1} seem to be useful for deciding whether a data set is of the AR or MA type. In the paper we use terminology like stationary time series, snapshot, filter, and linear transformation. This terminology is summarized in Fig. 1. The figure is referred to throughout the text.

Fig. 1. LIN FILT & LIN TRAN REPR OF RANDOM DATA



AR REPRESENTATIONS OF A STATIONARY TIME SERIES

In this section we briefly recall the results developed in [1] and provide an alternative interpretation of the maximum likelihood method spectrum.

Let (y_t) denote a real, zero-mean wide sense stationary sequence with real l_2 correlation sequence (r_t).

Kolmogorov Representation:

The sequence (y_t) has the following AR (∞) representation

$$\sum_{n=0}^{\infty} a_n y_{t-n} = u_t \qquad a_0 \neq 0$$

where u_t is a white noise sequence with zero mean and unit variance. The coefficients (a_i) are the AR (∞) filter coefficients.

The first order descriptor is readily obtained by replacing the input sequence (u_t) by (δ_t), $(\delta_{t=o}\ t \neq o,\ \delta_o=1)$:

$$\sum_{n=o}^{\infty} a_n h_{t-n} = \delta_t \qquad \text{for all } t$$

The correlation sequence, a second order descriptor, is characterized by

$$\sum_{m=o}^{\infty} a_m \sum_{n=o}^{\infty} r_{m-n+s}\, a_n = \delta_s$$

The specialization to the AR(p) case is straightforward.

AR or Analysis Transformations

We consider now a finite length data record $Y_t=(y_o,\ldots,y_{t-1})$ of the process (y_t). It has a symmetric and Toeplitz correlation matrix R_t:

$$R_t = \begin{bmatrix} r_o & r_1 & & r_{t-1} \\ r_1 & \ddots & \ddots & \\ & \ddots & \ddots & r_1 \\ r_{t-1} & & r_1 & r_o \end{bmatrix}$$

The inverse correlation matrix R_t^{-1} has a UL Cholesky decomposition (Fig. 1):

$$R_t^{-1} = A_t^H A_t$$

$$A_t = \begin{bmatrix} a_o^o & & \underline{a}_s & & 0 \\ & \ddots & & & \\ a_s^s & & a_o^s & & \\ & & & \ddots & \\ a_{t-1}^{t-1} & & a_{t-1-s}^{t-1} & & a_o^{t-1} \end{bmatrix} \quad \underline{a}^s$$

Defining the white vector U_t as $E(U_t)=o$ and $E(U_t U_t^H) = I_t$ (txt identity), the snapshot Y_t has the AR type representation (Fig.1).

$$A_t Y_t = U_t$$

The row vectors $\underline{a}^s$ of A_t are the order-s MA whiteners or analyzers and the column vectors $\underline{a}_s$ are interpreted as impulsive exciters [1]. $\underline{a}_s$ is also known as the impulse response for the AR filter that corresponds to the order-s MA synthesizer $\underline{h}^s$.

The first order descriptor is obtained by substituting I_t for U_t to obtain

$$A_t H_t = I_t$$

where H_t is the matrix

$$H_t = \begin{bmatrix} h_o^o & & \underline{h}_s & & 0 \\ & \ddots & & & \\ h_s^s & & h_o^s & & \\ & & & \ddots & \\ h_{t-1}^{t-1} & & h_{t-1-s}^{t-1} & & h_o^{t-1} \end{bmatrix} \quad \underline{h}^s$$

As $A_t Y_t = U_t$ describes a whitening operation on Y_t, replacing Y_t by $C_t(\omega)$ yields

$$A_t(\omega) = A_t C_t(\omega)$$

$$C_t^H(\omega) = (\,1\ e^{+j\omega}\ \ldots\ .\ e^{j\omega(t-1)}\,)$$

This is a column vector of phased complex frequency responses for the MA whiteners $\underline{a}^s$:

$$A_t(\omega) = (\,a^o(\omega),\ \ldots,\ a^{t-1}(\omega)\,)$$

$$a^s(\omega) = \underline{a}^s\, C_t(\omega)$$

The spectrum associated with Y_t is the inverse of the spectrum associated with the whitening transformation A_t (Norm of the complex frequency response):

$$R_t(\omega) = \frac{1}{|A_t(\omega)|^2} = \frac{1}{C_t^H(\omega) R_t^{-1} C_t(\omega)}$$

$$= \frac{1}{\sum_{s=o}^{t-1} |\,a^s(\omega)\,|^2}$$

This spectrum has the following interpretations:

- $R_t(\omega)$ is the maximum likelihood method (MLM) spectrum [3].
- The inverse of $R_t(\omega)$ is an average of order increasing whitening spectra [5].
- If only the (t-1) term is used in the summation above (corresponding to a very specific weighting pattern), the maximum entropy method (MEM) spectrum is produced [4].
- Using the second interpretation of the column vector $\underline{a}_s$, the MLM spectrum is also the Fourier transform of the diagonal sums of a correlation sequence. This correlation is built on the impulse response sequence for the AR filters that correspond to the order-s MA synthesizers $(o \leq s \leq t-1)$.

Parameterized AR Transformation

If the process (y_t) is AR (p), then in the Levinson/Durbin algorithm, the reflection coefficients converge to 0 in exactly p steps. Hence for t>p, the $\underline{a}^t$ vectors repeat themselves and the AR transformation specializes as follows:

$$A_{t,p} = \begin{bmatrix} a_o^o & & & & \\ & \ddots & & & 0 \\ a_p^p & & a_o^p & & \\ & \ddots & & \ddots & \\ 0 & & a_p^p & & a_o^p \end{bmatrix}$$

The results for AR(p) linear transformations suggest that the AR transformation A_t may be approximated by the p^{th} order approximant $A_{t,p}$ [1], [2]. The associated spectrum is a parameterized maximum likelihood spectrum. Its properties have been explored in [1], [2], [9].

MA REPRESENTATIONS OF A STATIONARY TIME SERIES

The same ideas translate directly to the MA case where the LU Cholesky decomposition of the correlation matrix R plays a key role (Fig.1). This yields an intermediate spectrum called the parameterized Bartlett spectrum (pBA).

Wold Decomposition

The Wold decomposition for the sequence (y_t) is the infinite moving average (MA(∞)):

$$y_t = \sum_{n=o}^{\infty} h_n u_{t-n}$$

The coefficients (h_n) are the MA(∞) filter coefficients. They are also the impulse response sequence since substituting (δ_t) for (u_t) yields:

$$h_t = h_t \qquad t > o$$

$$h_t = 0 \qquad ow$$

The correlation sequence (second order descriptor) is then written in terms of the first order descriptor:

$$r_t = \sum_{n=o}^{\infty} h_n h_{n+|t|} \qquad \text{for all } t$$

MA or Synthesis Transformation

As in the AR case, we consider only the snapshot Y_t. The corresponding R_t, symmetric and Toeplitz correlation matrix has the LU Cholesky decomposition

$$R_t = H_t H_t^H$$

where H_t is the lower triangular matrix introduced in the second section. The s^{th} column vector $\underline{h}^s$ is the order-s MA synthesizer and the s^{th} column vector $\underline{h}_s$ is the impulse response of an MA linear transformation to an impulse applied at time s. $\underline{h}_s$ is also the impulse response for the AR filter corresponding to the order-s MA whitener.

The finite length data record Y_t has the MA type representation

$$Y_t = H_t U_t$$

and the first order descriptor is obtained by replacing U_t by I_t: $H_t = H_t$. The frequency response of the MA transformations is

$$H_t(\omega) = C_t^H(\omega) H_t$$

This is a row vector of phased complex frequency responses for the impulse responses $(h_o(\omega), \ldots h_{t-1}(\omega)$. As in the AR linear transformation case, we define a spectrum

$$R'_t(\omega) = |H_t(\omega)|^2 = C_t^H(\omega) R_t C_t(\omega) = \left| \sum_{s=o}^{t-1} h_s(\omega) \right|^2$$

The spectrum $R'_t(\omega)$ has then the following interpretations:

- $\frac{1}{t} R'_t(\omega) = \sum_{n=-(t-1)}^{t-1} (1 - \frac{|n|}{t}) r_n e^{-jn\omega}$

- This is the conventional or Bartlett spectrum [7].

- It is also an average of magnitude squared frequency responses,

$$\frac{1}{t} R'_t(\omega) = \frac{1}{t} \sum_{s=o}^{t-1} |h_s(\omega)|^2$$

This is the reason for the extreme smoothness of the conventional spectrum.

- It is also interpreted in terms of the impulse responses for the AR filters corresponding to the order-s MA whiteners ($o \le s \le t-1$). It is the Fourier transform of the diagonal sums of the correlations built on these impulse responses.

Parameterized MA Transformations

As in the AR case, the parameterization is suggested by the particular structure of the matrix H_t in the MA(q) case. Suppose the original process (y_t) is MA(q). By running the impulse response algorithm we can generate the column vectors of H_t. The algorithm in that case produces time varying impulse responses of maximum length q. Hence the MA transformation specializes as follows:

$$H_t = \begin{bmatrix} h_o^o & & & 0 \\ & \ddots & & \\ h_q^q & \cdots & h_o^q & \\ & \ddots & & \ddots \\ 0 & & h_q^{t-1} & \cdots & h_o^{t-1} \end{bmatrix}$$

The order-s MA synthesizers h^s stop their growth at s=p but remain time varying . $\underline{h}^s$ converges to the true stationary MA(q) synthesizer only for $s \to \infty$. This is a major difference with the specialization of the AR transformation for the AR(p) case.

The parameterization suggested corresponds to applying a rectangular window on the correlation sequence and then using the impulse response algorithm. This procedure can lead to negative spectra that are meaningless, even though the Toeplitz nature of the approximant is preserved. We propose the following approximation:

$$H_{t,q} = \begin{bmatrix} h_o^o & & & 0 \\ & \ddots & & \\ h_q^q & \cdots & h_o^q & \\ & \ddots & & \ddots \\ 0 & & h_q^q & \cdots & h_o^q \end{bmatrix}$$

This approximation of H_t does not preserve the Toeplitz property but insures against negative spectra. As in the AR case, it corresponds to applying more weight to the q-order MA synthesizer. The spectrum associated with $H_{t,q}$ is a parameterized Bartlett (BA) spectrum:

$$\frac{1}{t} C_t^H(\omega) H_{t,q} H_{t,q}^H C_t(\omega)$$

NUMERICAL RESULTS

The influence of parameterization on the MLM and BA spectra is investigated for three different sets of synthetic data. For each set of data the correlation sequence is estimated using the unbiased estimator on 4096 points. The different spectra are calculated using 21 lags of the correlation sequence. The exact spectrum and the MEM spectrum are plotted as references.

The first set of data considered is the AR(4) process used in [1]. It has two real poles at $\pm$ .94 and a pair of complex conjugate poles at $\pm$ j .94. Fig. 2 shows the classical spectra corresponding to a covariance matrix of order 21. Fig. 3 shows what happens when the MLM and BA spectra are parameterized at (4,21). It is readily noted that the parameterization of the BA spectrum changes a great deal, whereas the MLM spectrum appears almost unaffected.

The second set of data is an MA(4) with two real zeros at $\pm$ .94 and a pair of complex conjugate zeros at $\pm$ j .94. Figs. 4-5 show the different spectra for the same parameterization as before. The MEM spectrum of order 21 has spurious peaks. Fig. 5 shows the dramatic changes obtained by parameterization of the MLM spectrum. Little effect is seen in the parameterized BA spectrum.

Finally the same experiment is carried out for a set of ARMA (2,1) data with a pair of complex conjugate poles at .66$\pm$j.49 and a real zero at .33. Here again we note the effect of parameterization. Both the BA and the MLM are affected, but the effect on the BA spectrum seems qualitatively much more important. For the MLM, the effects are evident only the low frequencies.

We can summarize the results in the following way. The selection of the right technique in parametric spectrum analysis depends a great deal on the type of data being analyzed. An MA type technique will be much more suited to the analysis of data of the MA type than an AR technique. This is obvious. Furthermore, once the right technique has been selected it is seen that parameterization has little effect on the resulting spectrum. These results seem to provide a method to differentiate between data sets of the MA, AR or ARMA type. One has to compare the parameterized MLM and Bartlett spectra to the non-parameterized ones to decide which one has been affected most.

CONCLUSIONS

We have presented a general framwork for deriving and interpreting analysis and synthesis spectra of the AR and MA type. Along the lines of [1] and [2] we have introduced a parameterized Bartlett spectrum. Here we haven't been concerned with the very important problem of order fitting. An order fitting rule (J-Divergence) was proposed in [1]. Discussions of MA(q) and AR(p) linear transformation generalize to ARMA(p,q) transformations. Such transformations underly lattice and innovations representations of stationary time series and are the object of future work.

REFERENCES

(1) L.L. Scharf, C.J. Gueguen, J.P. Dugre, "Parametric Spectrum Modelling: A Signal Processing Perspective," 1st IEEE Workshop on Spectrum Analysis, Hamilton, Aug. 1981.

(2) C.J. Gueguen, F. Gianella, "Analyse Spectrale et Approximation de Formes Quadratiques" Proc GRETSI, Nice, June 1981.

(3) R.T. Lacoss, "Data Adaptive Spectral Analysis Methods," Geophysics, Vol 26, 4, Aug. 1971.

(4) J.P. Burg, "Maximum Entropy Spectral Analysis," Ph.D. dissertation, Stanford University, 1975.

(5) J.P. Burg, "The Relationship Between Maximum Entropy Spectra and Maximum Likelihood Spectra," Geophysics, Vol. 37, Ap.1972.

(6) T.J. Ulrych, T.N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," Rev. Geophysics, Space Phys., Vol. 13, Feb. 1973.

(7) P.D. Welch, "The Use of the Fast Fourier Transform for the Estimation of Power Spectra," IEEE Trans. Audio Electroacoust, Vol.AU 15, June 1970.

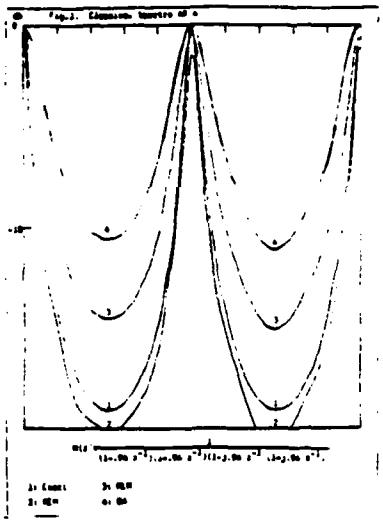
(8) C.J. Gueguen, L.L. Scharf, "Exact Likelihood Identification of ARMA Models: A Signal Processing Perspective," Proc. EUSIPCO, Lausanne, Switzerland, Sept. 1980.

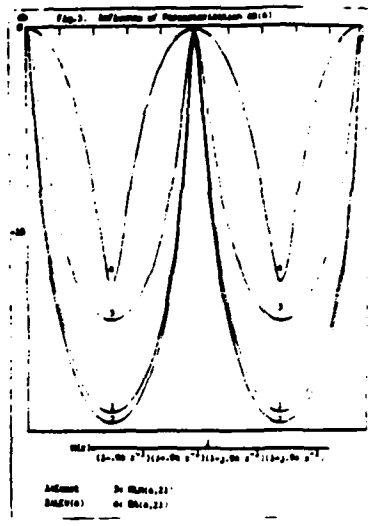
(9) J.P. Dugre, "Parametric Spectrum Analysis of Stationary Random Sequences," Ph.D. dissertation, Colorado State University, Aug. 1981.

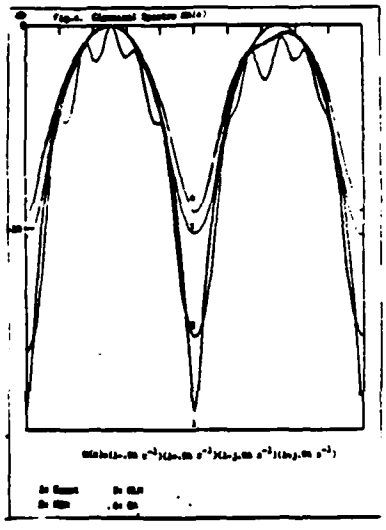
ACKNOWLEDGMENTS

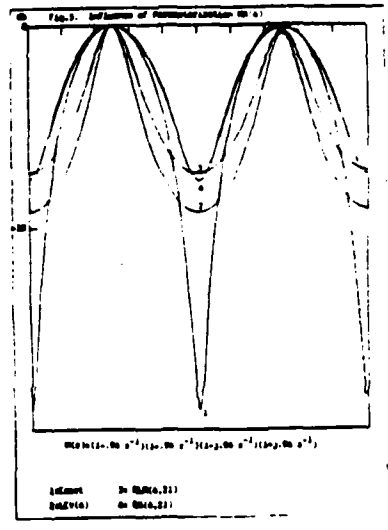
Our thinking has been influenced by recent results of M. Morf, T. Kailath, B. Friedlander and by remarks of A. Oppenheim at L'Aquila in September 1980.

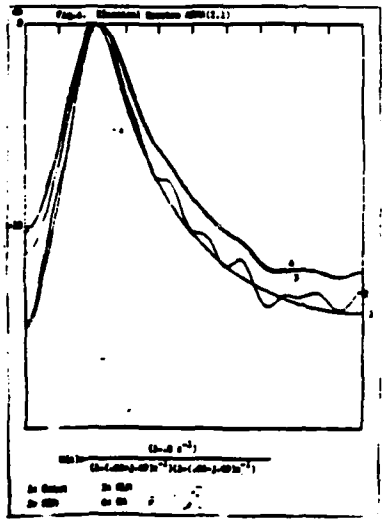
Continue to next page for Figures 2 through 7

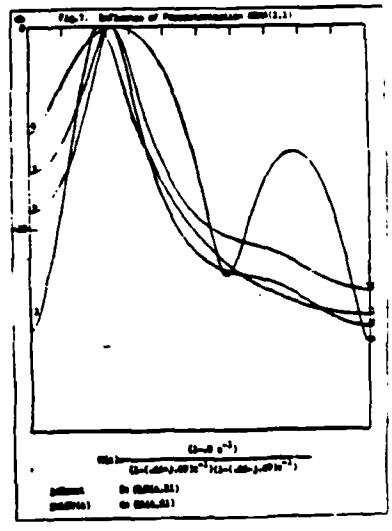












OFFICE OF NAVAL RESEARCH
STATISTICS AND PROBABILITY PROGRAM

BASIC DISTRIBUTION LIST
FOR
UNCLASSIFIED TECHNICAL REPORTS

FEBRUARY 1982

	Copies
Statistics and Probability Program (Code 411(SP)) Office of Naval Research Arlington, VA 22217	3
Defense Technical Information Center Cameron Station Alexandria, VA 22314	12
Commanding Officer Office of Naval Research Eastern/Central Regional Office Attn: Director for Science Barnes Building 495 Summer Street Boston, MA 02210	1
Commanding Officer Office of Naval Research Western Regional Office Attn: Dr. Richard Lau 1030 East Green Street Pasadena, CA 91101	1
U. S. ONR Liaison Office - Far East Attn: Scientific Director APO San Francisco 96503	1
Applied Mathematics Laboratory David Taylor Naval Ship Research and Development Center Attn: Mr. G. H. Gleissner Bethesda, Maryland 20084	1
Commandant of the Marine Coprs (Code AX) Attn: Dr. A. L. Slafkosky Scientific Advisor Washington, DC 20380	1
Navy Library National Space Technology Laboratory Attn: Navy Librarian Bay St. Louis, MS 39522	1
U. S. Army Research Office P.O. Box 12211 Attn: Dr. J. Chandra Research Triangle Park, NC 27706	1
Director National Security Agency Attn: R51, Dr. Maar Fort Meade, MD 20755	1
ATAA-SL, Library U.S. Army TRADOC Systems Analysis Activity Department of the Army White Sands Missile Range, NM 88002	1
ARI Field Unit-USAREUR Attn: Library c/o ODCSPER HQ USAEREUR & 7th Army APO New York 09403	1
Library, Code 1424 Naval Postgraduate School Monterey, CA 93940	1
Technical Information Division Naval Research Laboratory Washington, DC 20375	1
OASD (I&L), Pentagon Attn: Mr. Charles S. Smith Washington, DC 20301	1

Copies

Director
AMSAA
Attn: DRXSY-MP, H. Cohen
Aberdeen Proving Ground, MD 21005 1

Dr. Gerhard Heiche
Naval Air Systems Command (NAIR 03)
Jefferson Plaza No. 1
Arlington, VA 20360 1

Dr. Barbara Bailar
Associate Director, Statistical Standards
Bureau of Census
Washington, DC 20233 1

Leon Slavin
Naval Sea Systems Command (NSEA 05H)
Crystal Mall #4, Rm. 129
Washington, DC 20036 1

B. E. Clark
RR #2, Box 647-B
Graham, NC 27253 1

Naval Underwater Systems Center
Attn: Dr. Derrill J. Bordelon
Code 601
Newport, Rhode Island 02840 1

Naval Coastal Systems Center
Code 741
Attn: Mr. C. M. Bennett
Panama City, FL 32401 1

Naval Electronic Systems Command (NELEX 612)
Attn: John Schuster
National Center No. 1
Arlington, VA 20360 1

Defense Logistics Studies Information Exchange
Army Logistics Management Center
Attn: Mr. J. Dowling
Fort Lee, VA 23801 1

Copies

Reliability Analysis Center (RAC)
RADC/RBRAC
Attn: I. L. Krulac
Data Coordinator/
Government Programs
Griffiss AFB, New York 13441 1

Technical Library
Naval Ordnance Station
Indian Head, MD 20640 1

Library
Naval Ocean Systems Center
San Diego, CA 92152 1

Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20370 1

Mr. Dan Leonard
Code 8105
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Alan F. Petty
Code 7930
Naval Research Laboratory
Washington, DC 20375 1

Dr. M. J. Fischer
Defense Communications Agency
Defense Communications Engineering Center
1860 Wiehle Avenue
Reston, VA 22090 1

Mr. Jim Gates
Code 9211
Fleet Material Support Office
U. S. Navy Supply Center
Mechanicsburg, PA 17055 1

Mr. Ted Tupper
Code M-311C
Military Sealift Command
Department of the Navy
Washington, DC 20390 1

	Copies		Copies
Mr. F. R. Del Priori Code 224 Operational Test and Evaluation Force (OPTEVFOR) Norfolk, VA 23511	1		

END

FILMED

7-83

DTIC